Paidós
Básica

Daniel Dennett

La conciencia explicada

76

Una teoría interdisciplinar

La conciencia explicada

Daniel Dennett

La conciencia explicada

Una teoría interdisciplinar

Ilustraciones de Paul Weiner

ediciones
PAIDOS
Barcelona
Buenos Aires
México

Título original: *Consciousness explained*
Publicado en inglés por Little, Brown and Company

Traducción de Sergio Balari Ravera

Cubierta de Mario Eskenazi

1.ª edición, 1995

Quedan rigurosamente prohibidas, sin la autorización escrita de los titulares del "Copyright", bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler o préstamo públicos.

© 1991 by Daniel C. Dennett
© de todas las ediciones en castellano,
Ediciones Paidós Ibérica, S.A.,
Mariano Cubí, 92 - 08021 Barcelona,
y Editorial Paidós, SAICF,
Defensa, 599 - Buenos Aires.

ISBN: 84-493-0170-X
Depósito legal: B-24.469/1995

Impreso en Hurope, S.L.
Recaredo, 2 - 08005 Barcelona

Impreso en España - Printed in Spain

Para Nick, Marcel y Ray

SUMARIO

Prefacio 13

1. Preludio: ¿por qué son posibles las alucinaciones? 15
   1. Un cerebro en un tarro 15
   2. Bromistas en el cerebro 19
   3. Un juego de sociedad llamado psicoanálisis 22
   4. Presentación 28

Primera parte
PROBLEMAS Y MÉTODOS

2. Explicar la conciencia 33
   1. La caja de Pandora: ¿es necesario desmitificar la conciencia? 33
   2. El misterio de la conciencia 37
   3. Los atractivos de la sustancia mental 39
   4. Por qué el dualismo es un proyecto estéril 45
   5. El reto 51

3. Una visita al jardín fenomenológico 55
   1. Bienvenidos al fenome 55
   2. Nuestra experiencia del mundo exterior 58
   3. Nuestra experiencia del mundo interior 67
   4. Afecto 72

4. Un método para la fenomenología 79
   1. La primera persona del plural 79
   2. La perspectiva de la tercera persona 82
   3. El método de la heterofenomenología 84
   4. Mundos ficticios y mundos heterofenomenológicos 91
   5. El discreto encanto de la antropología 94
   6. A la descubierta de lo que uno está diciendo realmente 96
   7. Las imágenes mentales de Shakey 99
   8. La neutralidad de la heterofenomenología 107

Segunda parte
UNA TEORÍA EMPÍRICA DE LA MENTE

5. Versiones Múltiples frente al Teatro Cartesiano 115
1. El punto de vista del observador 115
2. Presentación del modelo de Versiones Múltiples 125
3. Revisiones orwellianas y estalinianas 129
4. Retorno al teatro de la conciencia 140
5. El modelo de Versiones Múltiples en acción 148

6. Tiempo y experiencia 153
1. Momentos efímeros y conejos saltarines 154
2. Cómo representa el tiempo el cerebro 158
3. Libet y el caso de la «referencia hacia atrás en el tiempo» .. 167
4. Las afirmaciones de Libet sobre el retraso subjetivo de la conciencia de las intenciones 176
5. Un regalo: el carrusel precognitivo de Grey Walter 180
6. Cabos sueltos 182

7. La evolución de la conciencia 185
1. En la caja negra de la conciencia 185
2. El principio 187
Escena primera: el nacimiento de los límites y de las razones 187
Escena segunda: nuevas y mejores maneras de producir el futuro 190
3. Evolución en los cerebros y el efecto Baldwin 195
4. Plasticidad en el cerebro humano: montando el decorado ... 200
5. La invención de buenos y malos hábitos de autoestimulación 207
6. El tercer proceso evolutivo: memas y evolución cultural 212
7. Los memas de la conciencia: la máquina virtual a instalar .. 222

8. Cómo las palabras hacen cosas con nosotros 241
1. Repaso: *¿e pluribus unum?* 241
2. Burocracia frente a pandemónium 245
3. Cuando las palabras quieren ser dichas 255

9. La arquitectura de la mente humana 267
1. ¿Dónde estamos? 267
2. Orientándonos con el pequeño esbozo 270
3. ¿Y ahora qué? 277
4. Los poderes de la máquina joyceana 288
5. Pero, ¿es esto una teoría de la conciencia? 294

Tercera parte
LOS PROBLEMAS FILOSÓFICOS DE LA CONCIENCIA

10. Mostrar y contar 299
1. Rotando imágenes en el ojo de la mente 299
2. Palabras, imágenes y pensamientos 311
3. Referir y expresar 316
4. Zombíes, zimbos y la ilusión del usuario 322
5. Problemas con la psicología del sentido común 327

11. Desmantelando el programa de protección de testigos 333
1. Resumen 333
2. Pacientes con visión ciega: ¿zombíes parciales? 334
3. Esconde el dedal: un ejercicio para despertar la conciencia 334
4. Visión protésica: ¿qué falta, además de la información? ... 349
5. «Repleción» frente a averiguación 355
6. La negligencia como pérdida patológica del apetito epistémico 367
7. Presencia virtual 370
8. Ver es creer: un diálogo con Otto 373

12. Los qualia descalificados 381
1. Una nueva cuerda para la cometa 381
2. ¿Por qué existen los colores? 386
3. Disfrutando de nuestras experiencias 395
4. Una fantasía filosófica: qualia invertidos 401
5. ¿Qualia «epifenoménicos»? 409
6. De vuelta a mi mecedora 417

13. La realidad de los yos 423
1. Cómo tejen los humanos un yo 424
2. ¿Cuántos yos por cliente? 429
3. La insoportable levedad del ser 436

14. La conciencia imaginada 443
1. Imaginando un robot consciente 443
2. Lo que se siente al ser un murciélago 452
3. Preocupación e importancia 460
4. La conciencia explicada, ¿o eliminada? 465

Apéndice A (Para filósofos) 467
Apéndice B (Para científicos) 473
Bibliografía 479
Índice analítico 499

PREFACIO

Durante mi primer curso de universidad leí las *Meditaciones* de Descartes y quedé prendado por el problema de la mente y el cuerpo. Ahí había un misterio. ¿Cómo es posible que mis pensamientos y mis sentimientos quepan en el mismo mundo que las células nerviosas y las moléculas que componen mi cerebro? Ahora, después de treinta años de pensar, hablar y escribir sobre este misterio, creo que he hecho algunos progresos. Creo que puedo trazar el perfil de una solución, una teoría de la conciencia que responde (o muestra cómo pueden hallarse respuestas) a las preguntas que han desconcertado tanto a filósofos y científicos como a profanos. Me han ayudado mucho. Ha sido una gran suerte el recibir las enseñanzas que, informalmente, infatigablemente e imperturbablemente, me transmitieron algunos pensadores admirables con los que nos iremos encontrando en estas páginas. Porque lo que aquí quiero contar no es la historia de una cogitación solitaria, sino la de una odisea a través de muchos mares, y las soluciones a los enigmas están fuertemente entrelazadas en una tela de diálogo y desacuerdo, donde a menudo aprendemos más de temerarios errores que de cautos equívocos. Estoy convencido de que todavía permanecen muchos errores en la teoría que aquí presentaré, y espero que sean de los peores, porque así otros propondrán mejores respuestas.

Las ideas presentadas en este libro han ido tomando forma a lo largo de muchos años, pero no inicié su redacción hasta enero de 1990, para terminar un año después gracias a la generosidad de varias instituciones y a la ayuda de muchos amigos, estudiantes y colegas. El Zentrum für Interdisziplinäre Forschung de Bielefeld, el CREA de la École Polytechnique de París y la Villa Serbelloni de la Rockefeller Foundation en Bellagio me proporcionaron las condiciones ideales para escribir y conferenciar durante los cinco primeros meses. Mi propia universidad, Tufts, ha apoyado mi trabajo a través del Center for Cognitive Studies, y me permitió presentar el penúltimo borrador en otoño de 1990 durante un seminario que reunió al profesorado y a los estudiantes de Tufts y otros buenos centros del área metropolitana de Boston. Quisiera expresar también mi agradecimiento a la Kapor Foundation y a la Harkness Foundation por su apoyo a nuestras investigaciones en el Center for Cognitive Studies.

Hace ya algunos años, Nicholas Humphrey acudió al Center for Cognitive Studies para trabajar conmigo; pronto, él, Ray Jackendoff, Marcel Kinsbourne y yo empezamos a reunirnos regularmente para debatir diversos aspectos y problemas relacionados con la conciencia. Sería difícil encontrar cuatro maneras tan diferentes de enfocar el problema de la mente, pero nues-

tras discusiones fueron tan provechosas y estimulantes que quiero dedicar el libro a estos cuatro buenos amigos en agradecimiento por todo lo que aprendí con ellos. Otros dos viejos colegas y amigos, Kathleen Akins y Bo Dahlbom, han jugado un papel fundamental en el desarrollo de mi pensamiento, por lo cual les estaré eternamente agradecido.

También quiero expresar mi gratitud al grupo ZIF de Bielefeld y, en particular, a Peter Bieri, Jaegwon Kim, David Rosenthal, Jay Rosenberg, Eckart Scheerer, Bob van Gulick, Hans Flohr y Lex van der Heiden; al grupo CREA de París, particularmente a Daniel Andler, Pierre Jacob, Francisco Varela, Dan Sperber y Deirdre Wilson; y a los «príncipes de la conciencia», que se unieron al grupo formado por Nick, Marcel, Ray y yo mismo en la Villa Serbelloni durante una intensa y productiva semana de marzo: Edoardo Bisiach, Bill Calvin, Tony Marcel y Aaron Sloman. Gracias también a Edoardo y a los demás participantes en el seminario sobre el abandono, celebrado en Parma en junio. También Pim Levelt, Odmar Neumann, Marvin Minsky, Oliver Selfridge y Nils Nilsson me proporcionaron sus valiosos consejos sobre el contenido de algunos capítulos. Mi agradecimiento también para Nils por aportar la fotografía de Shakey y a Paul Bach-y-Rita por sus fotografías y su asesoramiento sobre los dispositivos protésicos de visión.

Debo agradecer a todos los participantes del seminario del pasado otoño el regalo de su crítica constructiva; nunca olvidaré a esta clase: David Hilbert, Krista Lawlor, David Joslin, Cynthia Schossberger, Luc Faucher, Steve Weinstein, Oakes Spalding, Mini Jaikumar, Leah Steinberg, Jane Anderson, Jim Beattie, Evan Thompson, Turhan Canli, Michael Anthony, Martina Roepke, Beth Sangree, Ned Block, Jeff McConnell, Bjorn Ramberg, Phil Holcomb, Steve White, Owen Flanagan y Andrew Woodfield. Semana tras semana, esta pandilla me mantuvo contra las cuerdas de la manera más constructiva. Durante la redacción final, Kathleen Akins, Bo Dahlbom, Doug Hofstadter y Sue Stafford me sugirieron numerosas mejoras. Paul Weiner convirtió mis garabatos en unas excelentes figuras y diagramas.

Kathryn Wynes y, más tarde, Anne van Voorhis llevaron a cabo un extraordinario trabajo impidiendo que yo y el propio Center nos hundiéramos durante los últimos años de actividad frenética; sin su eficiencia y capacidad de previsión, a este libro todavía le quedarían muchos años para ser completado. Por último, y más importante: todo mi amor y agradecimiento para Susan, Peter, Andrea, Marvin y Brandon, mi familia.

Tufts University
Enero de 1991

Capítulo 1

PRELUDIO: ¿POR QUÉ SON POSIBLES LAS ALUCINACIONES?

1. Un cerebro en un tarro

Imagine usted que un grupo de malvados científicos le ha extirpado el cerebro mientras dormía y lo han introducido en un tarro con todo lo necesario para mantenerlo con vida. Imagine, además, que, hecho esto, los malvados científicos se dedican a hacerle creer que usted no es solamente un cerebro en un tarro, sino que sigue en pie con su cuerpo, participando en las actividades propias del mundo real. Esta vieja parábola, la del cerebro en un tarro, es uno de esos experimentos mentales favoritos que muchos filósofos siempre llevan en su zurrón. Es la versión moderna de aquel demonio malvado de Descartes (1641),[1] un ilusionista imaginario empeñado en hacer lo imposible por distraer a Descartes ante cualquier situación, incluida su propia existencia. Sin embargo, como observó el propio Descartes, ni siquiera un malvado demonio con poderes infinitos sería capaz de hacerle creer en su existencia si esto no fuera cierto: *cogito ergo sum*, «pienso, luego existo». Hoy en día, los filósofos están menos preocupados por probar su existencia en tanto que entes pensantes (quizá porque consideran que Descartes resolvió el problema satisfactoriamente) y más ocupados en tratar de responder a la pregunta de qué conclusiones debemos extraer de nuestras experiencias sobre la propia naturaleza y sobre la naturaleza del mundo en que (aparentemente) vivimos. ¿Es posible que usted no sea más que un cerebro en un tarro? ¿Es posible que usted *siempre* haya sido un cerebro en un tarro? Y si así fuera, ¿sería usted capaz de llegar a concebir su situación (por no hablar de confirmarla)?

El caso del cerebro en un tarro es un modo bastante ingenioso de aproximarse a estos problemas; sin embargo, quisiera utilizar esta parábola con un propósito ligeramente distinto. Me servirá para poner de manifiesto algunos hechos bastante sorprendentes en relación a las alucinaciones, los cuales, a su vez, nos encaminarán hacia los prolegómenos de una teoría —una teoría empírica y científicamente respetable— de la conciencia humana. El experimento mental, en su versión estándar, presupone que los malvados científicos poseen todos los medios a su alcance para transmitir a las termina-

1. Las fechas entre paréntesis remiten a los trabajos citados en la bibliografía.

ciones nerviosas de los sentidos los estímulos adecuados a fin de que su engaño tenga éxito, un supuesto que, aun reconociendo las evidentes dificultades técnicas que supondría, los filósofos han considerado como algo «posible en principio». Deberíamos ser un poco más cautos con aquello que, en principio, parece posible. En principio, también sería posible construir una escalera de acero hasta la Luna, o escribir en orden alfabético todas las conversaciones inteligibles llevadas a cabo en inglés que contuvieran menos de mil palabras. Sin embargo, ninguna de estas dos cosas es, de hecho, ni remotamente posible, y, a veces, *una imposibilidad de hecho* es teóricamente más interesante que *una posibilidad en principio*, como enseguida veremos.

Detengámonos sólo un momento para pensar en lo desalentadora que puede resultar la tarea emprendida por nuestros científicos malvados. Imaginémosles procediendo poco a poco, empezando por las tareas más sencillas hasta llegar a problemas de más difícil solución. Comenzarían con un cerebro convenientemente reducido a un estado comatoso, al que se mantiene con vida pero que no recibe ningún estímulo a través de los nervios ópticos, los nervios auditivos, los nervios somatosensoriales ni ninguna otra de las vías aferentes o, de entrada, del cerebro. Se suele asumir que un cerebro en estas condiciones permanecería en estado comatoso para siempre, sin necesidad de morfina para mantenerlo dormido, aunque existen algunas experiencias que parecen demostrar que un despertar repentino es posible incluso en circunstancias tan horribles como éstas. No me parece muy arriesgado afirmar que se sentiría usted bastante angustiado, si llegara a despertarse en tal estado: ciego, sordo, completamente insensible, sin ningún sentido de la orientación de su cuerpo. Sin ánimo de aterrorizarle, pues, los científicos deciden despertarle canalizando música en estéreo (debidamente codificada como impulsos nerviosos) hacia sus nervios auditivos. Producen también las señales apropiadas, que normalmente procederían de su sistema vestibular u oído interno, para hacerle creer que está usted tumbado boca arriba, aunque paralizado, insensible y ciego. Es muy probable que todo cuanto hemos descrito hasta ahora esté dentro de los límites del virtuosismo tecnológico en un futuro no muy lejano; quizá ya sea posible hoy en día. Nuestros científicos continuarían, entonces, con la estimulación de los canales que habían enervado su epidermis, comunicándoles lo que habría sido interpretado como una suave y uniforme sensación de calor sobre la superficie ventral de su cuerpo (la barriga), y (rizando el rizo) podrían estimular los nervios epidérmicos dorsales (posteriores) a fin de simular la hormigueante textura de finos granos de arena presionando sobre su espalda. «¡Estupendo!», pensaría usted, «aquí estoy, tumbado en la playa, paralizado y ciego, escuchando buena música, pero seguramente en peligro de quemarme al sol. ¿Cómo he llegado hasta aquí, y cómo puedo pedir ayuda?»

Supongamos ahora que los científicos, después de haber conseguido todo esto, se enfrentan al problema más complejo de convencerle de que usted no es una mera nuez de coco caída sobre la playa, sino un agente capaz de participar de diversas actividades en el mundo. Proceden paso a paso: deci-

den en primer lugar liberar parcialmente la «parálisis» de su cuerpo fantasma y le permiten mover el dedo índice de la mano derecha sobre la arena. De hecho, lo que hacen es transmitirle la experiencia sensorial del movimiento de su dedo, lo cual se consigue provocando la sensación de realimentación cinestésica* asociada a las señales motrices o volitivas relevantes en las terminaciones eferentes, o de salida, de su sistema nervioso. Pero también deben conseguir eliminar la insensibilidad de su dedo fantasma y producir los estímulos que provocarían la sensación de una arena imaginaria revuelta por efecto del movimiento del dedo.

De repente, nuestros científicos se ven enfrentados a un problema que pronto se les escapará de las manos, ya que la manera de percibir el movimiento de la arena depende de cómo decida usted mover el dedo. Calcular con propiedad la información necesaria para la realimentación, generarla y componerla, y finalmente presentarla en tiempo real se convertirá en un problema intratable computacionalmente, incluso en el más rápido de los ordenadores; alternativamente, si los malvados científicos deciden resolver el problema del tiempo real calculando previamente todas las respuestas posibles para poder «enlatarlas» y así reproducirlas cuando sea necesario, no conseguirán más que sustituir un problema insoluble por otro: hay demasiadas posibilidades que almacenar. En resumen, nuestros malvados científicos quedarán atrapados en el pantano de la *explosión combinatoria* en el mismo momento en que decidan concederle una mínima capacidad para explorar su mundo imaginario.[2]

Nuestros científicos han topado con un conocido obstáculo, cuya sombra se proyecta en los aburridos estereotipos de cualquier videojuego. Las alternativas abiertas para la acción deben quedar estrictamente —y en contra de todo realismo— limitadas a fin de que la labor de aquellos que quieren representar el mundo permanezca dentro de los límites de lo factible.

* Me permito aquí la libertad de acuñar el término (cinestesia), más todos sus derivados, para referirme al sentido por el cual percibimos el movimiento muscular, el peso, la posición, etc., de los miembros de nuestro cuerpo. La vieja idea aristotélica de que los sentidos son cinco (vista, oído, olfato, gusto y tacto) está totalmente superada; hoy en día, los especialistas conciben la piel humana como un órgano que participa en diversas modalidades de sentidos (por ej., calor, frío, presión y dolor); se reconocen, asimismo, otros sentidos como la cinestesia, que acabamos de definir, o el sentido del equilibrio. [N. del T.]

2. El término *explosión combinatoria* proviene de las ciencias de la computación y la informática, aunque el fenómeno es conocido desde mucho antes de la invención de los ordenadores, como queda demostrado en la fábula del emperador que accedió a recompensar a un campesino que le había salvado la vida con un grano de arroz en el primer cuadro de un tablero de ajedrez, dos granos en el segundo, cuatro en el tercero, y así sucesivamente, doblando la cantidad para cada uno de los sesenta y cuatro cuadros. El emperador acaba por adeudar al astuto campesino miles de millones de granos de arroz (2^{64}—1, para ser exactos). Más próximo a nuestro ejemplo es el caso de los novelistas «aleatorios» franceses que intentaron escribir novelas en las que, después de leer el capítulo 1, el lector debía lanzar una moneda para leer el capítulo 2a o 2b en función del resultado, repitiendo la operación para leer los capítulos 3aa, 3ab, 3ba o 3bb y así hasta terminar el libro. Estos novelistas pronto se percataron de que era necesario reducir el número de elecciones aleatorias, a fin de no provocar una explosión de ficción novelística que incluso impediría a los lectores transportar el «libro» desde la librería hasta su casa.

Si los científicos no pueden hacer otra cosa que convencerle de que está condenado a jugar a Donkey Kong toda la vida, entonces son realmente unos científicos perversos.

Para este problema técnico existe algo parecido a una solución. Es la solución utilizada, por ejemplo, para reducir la carga computacional en aquellos simuladores de vuelo que poseen un gran realismo: el uso de *réplicas* de los elementos del mundo simulado. Utilice una carlinga de verdad y muévala con elevadores hidráulicos en vez de intentar simular toda esta información con el asiento del piloto que se está entrenando. En resumen, sólo hay un modo de que usted pueda almacenar tanta información sobre un mundo *real* (quizá minúsculo, artificial, de escayola) para almacenar su propia información. Eso es «hacer trampas», particularmente si usted es el genio malvado que afirma haber engañado a Descartes sobre la existencia de absolutamente todo, pero es una manera de hacer el trabajo con algo menos que recursos infinitos.

Descartes tuvo el buen criterio de dotar a su genio imaginario con poderes *infinitos* para el engaño. Aunque la empresa no es, estrictamente hablando, infinita, la cantidad de información que en poco tiempo puede obtener un ser humano mínimamente inquisitivo es impresionante. Los ingenieros miden el flujo de información en *bits* por segundo, o hablan del *ancho de banda* de los canales a través de los cuales fluye la información. La televisión tiene un ancho de banda mayor que la radio, y la televisión de alta definición lo tiene aún mayor. La televisión de alta definición sensitiva tendría un ancho de banda todavía mayor, y la televisión sensitiva *interactiva* tendría un ancho de banda astronómico, ya que éste se ramificaría constantemente en miles de trayectorias ligeramente distintas a través del mundo (imaginario). Déle a un escéptico una moneda de dudoso valor y, después de un par de segundos sopesándola, rascándola, escuchando su tintineo, mordiéndola o simplemente observando cómo refleja la luz del Sol, el escéptico habrá consumido más *bits* de información de la que uno de los superordenadores Cray* puede organizar en un año. Fabricar una moneda falsa, pero *real*, es un juego de niños; producir una moneda *simulada* a partir tan sólo de una serie de estímulos nerviosos organizados supera las capacidades de la tecnología humana actual, y probablemente la de todos los tiempos.[3]

* Seymour R. Cray, ingeniero electrónico e informático, es el diseñador de los superordenadores Cray-1 y Cray-2, que son los más rápidos construidos hasta el momento. [N. del T.]

3. Hoy en día vivimos el *boom* del desarrollo de sistemas de «realidad virtual» para el ocio y la investigación. El nivel de sofisticación alcanzado es impresionante: con guantes electrónicos que constituyen una interficie bastante convincente para «manipular» objetos virtuales, y cascos-pantalla que permiten explorar espacios virtuales de una complejidad considerable. No obstante, las limitaciones de estos sistemas son aparentes y corroboran mis observaciones: la robustez de estas ilusiones sólo puede mantenerse gracias a la repetida combinación de réplicas físicas y a la esquematización (una representación *relativamente* tosca). Además, incluso en los mejores casos, son experiencias de subrealidad virtual y no algo que uno pueda confundir con la realidad por mucho tiempo. Si usted quiere hacerle creer a alguien que está encerrado en una jaula con un gorila, contratar los servicios de un actor disfrazado de gorila va a ser su apuesta más segura durante mucho tiempo.

Una de las conclusiones que podemos extraer de todo cuanto hemos dicho hasta ahora es que no somos cerebros en un tarro (esto, por si usted empezaba a estar preocupado). Otra conclusión que aparentemente podemos extraer es que las alucinaciones fuertes son imposibles, donde por alucinación fuerte entiendo la alucinación de un objeto aparentemente real, concreto y en tres dimensiones, en oposición a destellos, distorsiones geométricas, auras, imágenes accidentales, experiencias fugaces con extremidades fantasma y otras sensaciones anómalas. Una alucinación fuerte sería, por ejemplo, un fantasma que respondiera a nuestras palabras, que se dejara tocar y produjera una sensación de solidez, que proyectara una sombra, que fuera visible desde cualquier ángulo de modo que pudiéramos caminar a su alrededor para ver cómo es su espalda.

Es posible hacer una clasificación aproximada de las alucinaciones en función del número de rasgos de este tipo que presentan. Los testimonios de alucinaciones *muy* fuertes son raros, y ahora podemos comprender por qué no es una coincidencia que, intuitivamente, la credibilidad de tales testimonios sea inversamente proporcional a la fuerza de la alucinación descrita. Somos —y debemos ser— particularmente escépticos ante los testimonios de alucinaciones muy fuertes porque no creemos en fantasmas, y porque pensamos que sólo un verdadero fantasma puede producir una alucinación fuerte. (La verdadera fuerza de las alucinaciones relatadas por Carlos Castañeda en *Las enseñanzas de Don Juan: una forma yaqui de conocimiento* [1968] supuso para los científicos el principal indicio de que el libro, pese a haber sido una celebrada tesis doctoral de antropología en UCLA, tenía más de ficción que de hecho.)

No obstante, aunque no podamos afirmar que las alucinaciones *verdaderamente* fuertes son posibles, no cabe duda de que con frecuencia se experimentan convincentes alucinaciones multimodales. Las alucinaciones documentadas en la bibliografía de psicología clínica son a menudo detalladas fantasías que están muy por encima de la capacidad generativa de la tecnología actual. ¿Cómo es posible que un solo cerebro haga lo que una legión de científicos y expertos en animación por ordenador serían prácticamente incapaces de hacer? Si tales experiencias no son percepciones genuinas o verídicas de una cosa real «fuera» de la mente, entonces deben producirse enteramente dentro de la mente (o del cerebro); tramadas de manera artificial, pero lo suficientemente próximas a la realidad como para confundir a la propia mente que las inventó.

2. Bromistas en el cerebro

La explicación que más comúnmente se suele dar a este problema es la de suponer que las alucinaciones ocurren cuando se produce algún tipo de autoestimulación anormal del cerebro: una estimulación de ciertas partes o niveles de los sistemas perceptivos del cerebro generada internamente en

su totalidad. Descartes, en el siglo XVII, contempló claramente esta posibilidad en su análisis del fenómeno de las extremidades fantasma: la alucinación, sorprendente aunque bastante común, que experimentan los amputados cuando sienten no sólo la presencia del miembro amputado, sino también picores, hormigueos y dolores en esa parte del cuerpo. (Con frecuencia ocurre que los que acaban de ser sometidos a una intervención de este tipo no creen que la pierna o el pie han sido amputados, hasta que *ven* que efectivamente ya no están ahí; tan claras y realistas son las sensaciones de su continua presencia.) Descartes utilizó la campanilla como analogía. Antes de que hubiera timbres eléctricos, intercomunicadores y *walkie-talkies*, las grandes casas poseían unos complejos sistemas de cables y poleas que permitían avisar a los sirvientes desde cualquiera de las habitaciones de la casa. Un firme tirón del cordón de terciopelo que colgaba de un agujero en la pared era suficiente para tirar de un cable que, gracias a un sistema de poleas, estaba conectado a una campanilla numerada en las dependencias del servicio. Cada vez que una de esas campanillas sonaba, el mayordomo sabía que se requería de sus servicios en el dormitorio del señor, en el salón o en la sala de billar. Estos sistemas funcionaban muy bien, pero estaban hechos a la medida de cualquier bromista: un tirón al cable del salón en cualquier punto de su recorrido era suficiente para que el mayordomo corriera hacia allí, comprobara que la habitación estaba vacía y se quedara con la duda de si había sido llamado o no; una especie de pequeña alucinación. Similarmente, Descartes pensó que, dado que las percepciones son causadas por complicadas cadenas de eventos en el sistema nervioso que terminan en el centro de control de la mente consciente, si fuera posible intervenir en algún punto de esta cadena (por ejemplo, en algún punto del nervio óptico, entre el globo ocular y la conciencia), un tirón en el punto apropiado de los nervios produciría exactamente la cadena de eventos que causaría la percepción real de algo, lo cual, a su vez, produciría en el punto de recepción de la mente exactamente los mismos efectos que una percepción consciente.

El cerebro, o alguna de sus partes, en un descuido, engaña a la mente. Ésta fue la explicación que dio Descartes al fenómeno de los miembros fantasma. Sin embargo, las alucinaciones de miembros fantasma, pese a ser bastante vívidas, son, de acuerdo con nuestra terminología, relativamente débiles; consisten en una serie desordenada de dolores y picores, todos dentro de una única modalidad sensorial. Los amputados no ven, oyen ni (por lo que yo sé) huelen sus pies fantasma. Así pues, dejando de lado por el momento el enigma de la interacción entre el cerebro físico y la mente consciente no física, una explicación como la de Descartes *podría* ser la correcta para el fenómeno de los miembros fantasma. No obstante, podemos demostrar que ni siquiera la parte puramente mecánica del análisis de Descartes puede ser correcta como explicación de alucinaciones relativamente fuertes: no hay manera de que el cerebro ilusionista pueda almacenar y manipular la información falsa suficiente para engañar a una mente inquisitiva.

El cerebro puede relajarse y dejar que el mundo real proporcione grandes dosis de información *verdadera*, pero si empieza a intentar cortocircuitar sus propios nervios (o a tirar de sus propios cables, como habría dicho Descartes), el resultado no pasará de la más débil y efímera de las alucinaciones. (De igual modo, el mal funcionamiento del secador de pelo de un vecino puede causar «nieve» o «moscas», y zumbidos o raros destellos en la pantalla de nuestro televisor, pero si vemos una versión falsa del telediario del mediodía, *sabemos* que lo que la ha provocado tiene una causa mucho más compleja que está muy por encima de las capacidades de un secador de pelo.)

Resulta tentador suponer que quizá hemos sido un poco crédulos ante el caso de las alucinaciones. Quizá sólo las alucinaciones suaves, fugaces y débiles se producen alguna vez; las alucinaciones fuertes nunca se dan porque no pueden darse. Un somero repaso de la bibliografía sobre alucinaciones sugiere claramente que existe algo así como una relación inversa entre fuerza y frecuencia de las alucinaciones, así como entre fuerza y credibilidad. Tal repaso nos proporciona asimismo nuevas pistas para la elaboración de una teoría sobre los mecanismos de producción de alucinaciones. Una característica endémica de todo testimonio de alguna alucinación es el reconocimiento de una pasividad poco usual ante la alucinación por parte de aquellos que han sido víctimas de un fenómeno de este tipo: los alucinados se limitan a contemplar el fenómeno maravillados, pero jamás sienten el deseo de investigar o explorar, y nunca intentan interactuar con las apariciones. Es muy probable que tal pasividad sea, por los motivos que acabamos de exponer, una característica esencial de las alucinaciones, un requisito necesario para que se produzca una alucinación mínimamente detallada y duradera.

La pasividad, sin embargo, no es más que un caso especial de la manera en que una alucinación relativamente fuerte puede sobrevivir. El motivo por el cual tales alucinaciones pueden sobrevivir es que el ilusionista —palabra con la que quiero designar al responsable de que se produzcan alucinaciones, quienquiera que éste sea— puede «contar con» que la víctima será más o menos activa en el momento de investigar el fenómeno; en el caso de la pasividad total, la actividad investigadora será nula. En tanto en cuanto el ilusionista sea capaz de predecir con detalle el grado de actividad investigadora de la víctima, no tiene más que hacer que la ilusión se mantenga «en las diversas perspectivas desde las que mirará la víctima». Los diseñadores de decorados de cine siempre insisten en conocer de antemano la colocación exacta de la cámara o, si ésta no ha de permanecer estacionaria, su trayectoria y ángulo precisos. Así sólo tienen que preparar el material necesario para cubrir las perspectivas que entrarán dentro del encuadre. (No es una casualidad que en el *cinéma vérité* se haga un uso extensivo de tomas con la cámara al hombro.) En la vida real, Potemkin utilizó el mismo principio para decorar los pueblos que debía visitar Catalina la Grande: su itinerario tuvo que ser acorazado.

Así pues, una solución para el problema de las alucinaciones fuertes será

suponer que existe un vínculo entre la víctima y el ilusionista, lo cual permite a este último construir una ilusión que *depende* de la intención prospectiva y de las decisiones de la víctima, lo que presupone que es capaz de anticipar estas reacciones. Cuando el ilusionista no es capaz de «leer la mente de la víctima» para obtener la información necesaria, en la vida real siempre queda el recurso de que el ilusionista (un mago de feria, por ejemplo) polarice la atención en la dirección deseada mediante un sutil pero poderoso «golpe de efecto psicológico». Así, un mago que realiza un número con naipes utiliza diversas técnicas para crear en la víctima la ilusión de que elige libremente las cartas sobre el tapete, cuando en el fondo sólo hay una carta que puede ser vuelta boca arriba. Volviendo ahora a nuestro experimento mental, si los científicos malvados pudieran *obligar* al cerebro del tarro a tener unas determinadas intenciones prospectivas, podrían entonces salvar el problema de la explosión combinatoria preparando sólo el material necesario; el sistema sería interactivo sólo *en apariencia*. De igual modo, el genio perverso de Descartes podría obrar su propósito con poderes finitos, si fuera capaz de mantener la ilusión del libre albedrío de la víctima, cuya capacidad prospectiva del mundo imaginario es controlada por el propio genio malvado.[4]

Sin embargo, existe un modo aún más económico (y realista) de que las alucinaciones sean producidas en el cerebro, un sistema que aprovecha la propia curiosidad de la víctima. Podremos comprender mejor cómo funciona utilizando como analogía un conocido juego de sociedad.

3. Un juego de sociedad llamado psicoanálisis

En este juego se convence a una persona, el inocente del grupo, para que abandone la habitación al tiempo que se le informa de que otro de los miembros del grupo relatará algún sueño que haya tenido recientemente. Éste deberá contar su sueño a todos los que se queden en la habitación, de modo que cuando vuelva aquel a quien le tocó salir y empiece a hacer preguntas, la identidad del que tuvo el sueño quedará oculta entre el coro de los que responden. El cometido del que salió es hacer preguntas, a las que los demás sólo pueden responder sí o no, a fin de adivinar la trama del sueño con el mayor detalle posible y, una vez hecho esto, deberá psicoanalizar al que soñó para poder adivinar su identidad.

Una vez que uno de los asistentes ha abandonado la habitación, el huésped explica al resto del grupo que nadie debe contar un sueño, sino que deberán responder a las preguntas de acuerdo con las siguientes reglas: si la última letra de la última palabra de la pregunta está en la primera mitad

4. Para un análisis más detallado de los problemas del libre albedrío, el control, la lectura de la mente y la anticipación, véase mi libro *La libertad de acción. Un análisis de las exigencias del libre albedrío*, 1984a, especialmente los capítulos 3 y 4.

del alfabeto, tendrán que responder sí, y tendrán que responder no en los demás casos; sólo hay una condición: se aplica una regla especial para evitar contradicciones que invalida a las anteriores, de acuerdo con la cual debe responderse a preguntas posteriores de forma que no contradigan las respuestas de preguntas anteriores. Por ejemplo:

P: ¿Trata el sueño de una chica?
R: Sí.

Pero si más adelante se formula una pregunta como

P: ¿Aparecen personajes femeninos?
R: Sí [a pesar de la *s* final, en aplicación de la regla que prohíbe las contradicciones].[5]

Cuando el inocente vuelve a la habitación y empieza a hacer preguntas, recibe como respuesta una serie de síes y noes ordenados al azar, o en todo caso distribuidos arbitrariamente. Casi siempre el resultado es muy entretenido. A veces el juego termina de repente y de forma bastante absurda, como ocurriría si la primera pregunta fuera «¿es el sueño idéntico palabra por palabra a *El Quijote*?» o también «¿aparecen seres animados?». Aunque lo más normal es que, para regocijo de los participantes, se vaya construyendo una historia estrafalaria y a menudo obscena, plagada de ridículos percances. Cuando finalmente quien hace las preguntas decide que el que tuvo el sueño —quienquiera que éste o ésta sea— debe ser alguien muy enfermo y lleno de complejos, el animado grupo se apresura a informarle de que el autor del «sueño» no es otro más que él mismo. Evidentemente, esto no es del todo cierto. De alguna manera, el inocente sí que es el autor del sueño en virtud de las preguntas que decidió hacer. (Sólo *a él* se le ocurrió poner a los tres gorilas en el bote de remos con la monja.) Pero en otro sentido, el sueño simplemente carece de autor, y eso es precisamente lo importante. Estamos ante un proceso de producción narrativa, de acumulación de detalles, sin plan alguno ni intención de ser autor: una ilusión sin ilusionista.

La estructura de este juego tiene un sorprendente parecido con la estructura de una familia de modelos bien conocidos de los sistemas perceptivos. Es un hecho comúnmente aceptado que la visión humana, por ejemplo, no puede ser explicada como un proceso *únicamente* «dirigido por los datos» o «de abajo arriba», sino que es necesario suponer la existencia, en los niveles superiores, de ciertos procesos «dirigidos por expectativas» para la verificación de hipótesis (o algo parecido a la verificación de hipótesis). Otro miembro de esta familia de modelos es el modelo perceptivo de «análisis

5. Tras numerosos ensayos, se ha llegado a la conclusión de que el juego produce una historia mejor si se favorecen ligeramente las respuestas afirmativas, fijando en p/q la línea divisoria entre el sí y el no.

por síntesis» que también supone que las percepciones se construyen en un proceso que combina expectativas generadas centralmente, por un lado, y confirmaciones (y desmentidos) producidas en la periferia, por el otro (véase Neisser, 1967). La idea general que subyace en estas teorías es que una vez se ha producido una determinada cantidad de «preprocesamiento» en los estratos iniciales o periféricos del sistema perceptivo, las tareas perceptivas se completan —los objetos son identificados, reconocidos, categorizados— con una serie de ciclos de generación y verificación. En cada uno de esos ciclos, las expectativas e intereses del momento nos sirven para elaborar hipótesis que nuestros sistemas perceptivos deben confirmar o refutar; una rápida secuencia de generaciones y confirmaciones de hipótesis da lugar al producto final, el «modelo presente», puesto al día, del mundo del sujeto perceptor. La base sobre la que se fundamentan dichas explicaciones de la percepción responde a consideraciones de diversa índole, tanto biológicas como epistemológicas. No podemos decir que existan pruebas irrefutables en favor de tales modelos, si bien es cierto que los experimentos que se han llevado a cabo inspirados por este enfoque han tenido un éxito notable. Algunos teóricos han sido tan osados como para afirmar que la percepción *debe* tener esta estructura fundamental.

Sea cual sea el veredicto final en cuanto a la veracidad de las teorías de la percepción basadas en la generación y verificación de hipótesis, observamos que nos permiten dar una explicación simple y bastante sólida del fenómeno de las alucinaciones. Lo único que es preciso suponer para que un sistema perceptivo normal entre en un estado de alucinación es que, mientras la parte de generación de hipótesis del ciclo (es decir, la que está dirigida por expectativas) funciona con normalidad, la parte dirigida por los datos (es decir, la encargada de verificar) entra en un proceso aleatorio o arbitrario de confirmación y refutación, exactamente igual que en nuestro juego de sociedad. En otras palabras, si el ruido en el canal de datos se ve arbitrariamente amplificado en forma de «confirmaciones» y «refutaciones» (las respuestas arbitrarias en forma de sí o no del juego), las expectativas, inquietudes, obsesiones y preocupaciones que pueda tener la víctima en ese momento harán que se planteen preguntas o hipótesis cuyo contenido reflejará, con toda seguridad, esos intereses. De este modo, en el sistema perceptivo se irá desplegando una «historia» sin autor. No es necesario suponer que la historia estaba escrita de antemano, ni tampoco que la información se almacena y se combina en la parte ilusionista del cerebro. Lo único que hace falta asumir es que el ilusionista entra en un estado arbitrario de verificación y que la víctima proporciona el contenido al plantear sus preguntas.

Esta explicación es la que nos permite establecer un vínculo más directo entre el estado emocional del alucinado y el contenido de las alucinaciones que se producen. Por lo general, el contenido de las alucinaciones está relacionado con las inquietudes que asaltan al alucinado en un determinado momento, y el modelo que acabamos de exponer incorpora este factor sin necesidad de recurrir a la intervención de un narrador interno dotado de un grado

de información inverosímil y poseedor de una teoría o un modelo de la psicología de la víctima. Por ejemplo, ¿por qué un cazador, el último día antes de la veda, *ve* un ciervo, con sus astas y su blanca cola, cuando lo que está mirando es una vaca negra o a otro cazador con una chaqueta naranja? Porque su interrogador interno está preguntando obsesivamente «¿es un ciervo?» y recibiendo un NO como respuesta, hasta que finalmente un poco de ruido en el sistema se ve amplificado por error como un SÍ, con las consabidas consecuencias catastróficas.

Existe un cierto número de descubrimientos que avalan esta concepción de las alucinaciones. Por ejemplo, es bien sabido que las alucinaciones son el resultado normal de largos períodos de privación sensorial (véase, por ejemplo, Vosberg, Fraser y Guehl, 1960). Una posible explicación de este hecho sería que durante la privación sensorial, la parte dirigida por datos del sistema de generación y verificación de hipótesis haga descender la posición de su umbral para el ruido, el cual se ve así amplificado en forma de patrones arbitrarios de señales de confirmación y refutación, que terminan por convertirse en detalladas alucinaciones, cuyo contenido no es sino el producto de ansiosas expectativas y confirmaciones aleatorias. Además, como demuestra la mayoría de testimonios, las alucinaciones se van produciendo de forma gradual (en condiciones de privación sensorial o bajo el efecto de drogas). Primero son débiles —por ejemplo, geométricas—, para ir haciéndose más fuertes («objetivas» o «narrativas»); efecto, éste, que forma parte de las predicciones de nuestro modelo (véase, por ejemplo, Siegel y West, 1975).

Finalmente, el mero hecho de que, por difusión en el sistema nervioso, una droga sea capaz de producir efectos tan complejos y ricos en contenido también necesita de una explicación; es evidente que la droga *por sí misma* no puede «contener la historia», por mucho que algunos incautos se empeñen en creer que es así. Es muy poco probable que una droga, actuando de forma difusa, pueda crear o incluso convertirse en un complejo sistema de ilusionismo, mientras que es fácil ver que una droga podría actuar de manera que el umbral de verificación de un sistema de generación de hipótesis se viera rebajado, elevado o simplemente alterado de manera arbitraria.

Es evidente que el modelo de generación de alucinaciones inspirado en el juego de sociedad puede dar cuenta también de la composición de los sueños. Desde los trabajos de Freud, han quedado pocas dudas sobre el hecho de que el contenido temático de los sueños es claramente sintomático de los más profundos impulsos, ansiedades y preocupaciones del que sueña, aunque los indicios que nos proporcionan los sueños siempre quedan bien ocultos por un barniz de simbolismo y de pistas falsas. ¿Qué otro proceso podría producir historias que con tanta efectividad hablan sin cesar de las más profundas inquietudes del que sueña y a la vez ocultarlo todo bajo el velo de la metáfora y la sustitución? El freudiano suele responder a esta pregunta con la extravagante hipótesis de que existe un dramaturgo interno especializado en componer sueños terapéuticos en beneficio del ego y en evitar astutamente las intervenciones de un censor interno disfrazando su auténtico

significado. (El modelo freudiano podría denominarse también modelo de Hamlet, ya que recuerda la estrategia de Hamlet de ofrecer una representación de «La ratonera» sólo para Claudio; se necesita un genio realmente listo para imaginar una estratagema como ésta, pero si debemos creer a Freud, todos albergamos a un virtuoso de la narración de este tipo.) Como veremos más adelante, no todas aquellas teorías que proponen la existencia de *homúnculos* («hombrecillos» en el cerebro) merecen ser rechazadas; sin embargo, siempre que uno deba apelar a los homúnculos para que corran en su ayuda, sería conveniente que éstos se parecieran más a una brigada de funcionarios estúpidos que a los brillantes dramaturgos de Freud, encargados de organizarnos una representación onírica para cada noche. El modelo que hemos propuesto elimina la necesidad del dramaturgo por completo y cuenta con la «audiencia» (por analogía con aquel que «lo es» en el juego de sociedad) para que se encargue de aportar el contenido. Está claro que la audiencia no es un simple figurante, pero por lo menos no tiene por qué disponer de una teoría sobre sus propias ansiedades; sólo debe dejarse llevar por ellas en el momento de hacer las preguntas.

Merece la pena señalar asimismo, que un rasgo esencial del juego de sociedad como la regla que prohíbe las contradicciones no sería necesario en un proceso de producción de sueños o alucinaciones. Dado que los sistemas perceptivos siempre parecen estar explorando una situación que se está produciendo en ese momento (y no un *fait accompli*, la narración de un sueño ya contado), cualquier confirmación subsiguiente que pudiera resultar «contradictoria» puede ser interpretada por el mecanismo como la indicación de un cambio en el mundo en vez de como una revisión de la historia conocida por los que están contando el sueño. El fantasma era azul cuando miré por primera vez, pero ahora se ha vuelto verde; sus manos se convirtieron en garras, y así sucesivamente. La volatilidad de las metamorfosis de los objetos en los sueños y las alucinaciones es una de las características más sorprendentes de estos relatos y aún más sorprendente es lo poco que nos «preocupan» tales metamorfosis cuando soñamos. Así que la granja en Vermont se convierte de repente en un banco en Puerto Rico, y el caballo que estaba montando es ahora un coche, no... una lancha motora, y mi acompañante, que inició el viaje siendo mi abuela, es ahora el Papa. Estas cosas pasan.

Esta volatilidad es precisamente lo que uno esperaría de un inquisidor activo pero poco proclive al escepticismo, enfrentado a un conjunto aleatorio de síes y noes. Por otro lado, la persistencia de ciertos temas y objetos en los sueños, su resistencia a metamorfosearse o desaparecer, también puede ser explicada por nuestro modelo. Si aceptamos, por el momento, que el cerebro utiliza la regla del alfabeto y lleva a cabo el procesamiento en castellano, podemos imaginar cómo el siguiente interrogatorio subterráneo puede llegar a crear un sueño obsesivo:

P: ¿Trata de mi padre, quizás?
R: No.

P: ¿Trata de una llamada telefónica?
R: Sí.
P: Trata de mi madre, entonces.
R: No.
P: ¿No trata de mi padre, pues?
R: No.
P: ¿Es sobre mi padre que telefonea?
R: Sí.
P: ¡Ya sabía yo que era sobre mi padre! ¿Me está llamando a mí?
R: Sí...

Apenas podemos decir que la teoría que hemos esbozado demuestre algo (todavía) sobre los sueños y las alucinaciones. Demuestra, metafóricamente, cómo *podría* ser una explicación mecanicista de estos fenómenos, lo cual es bastante para un preludio, ya que muchos se sienten tentados por la tesis derrotista que sostiene que la ciencia no puede explicar «en principio» los diversos «misterios» de la mente. Lo presentado hasta ahora ni siquiera aborda el problema de nuestra *conciencia* de los sueños y las alucinaciones. Además, aunque hemos podido exorcizar a un homúnculo improbable, el inteligente ilusionista/dramaturgo que engaña a la mente, de momento lo hemos sustituido no sólo por unos contestadores de preguntas estúpidos (que bien podríamos sustituir a su vez por máquinas), sino también por el todavía demasiado listo, y poco explicado, inquisidor, la «audiencia». Quizá hayamos conseguido eliminar a un villano, pero aún no hemos empezado siquiera a ocuparnos de la víctima.

No podemos negar, no obstante, que hayamos hecho algunos progresos. Hemos visto que, al atender los requisitos «ingenieriles» de un fenómeno mental, surgen nuevas preguntas más fáciles de contestar, como por ejemplo: ¿qué modelos de las alucinaciones son capaces de evitar la explosión combinatoria? ¿De qué manera puede ser elaborado el contenido de la experiencia por unos procesos ciegos y (relativamente) estúpidos? ¿Qué tipo de vínculos entre procesos o sistemas podrían explicar el resultado de su interacción? Si queremos elaborar una teoría científica sobre la conciencia, deberemos enfrentarnos a muchas preguntas como éstas.

También hemos introducido una respuesta que será fundamental en todo lo que seguirá. El elemento clave de cuantas explicaciones hemos aventurado aquí sobre por qué son posibles los sueños y las alucinaciones es que lo único que debe hacer el cerebro es llevar a cabo todo lo que sea necesario para *aliviar el hambre epistémica*; satisfacer la «curiosidad» en todas sus formas. Si la «víctima» es pasiva o indiferente ante el asunto x, si la víctima no busca respuestas a ninguna pregunta sobre x, entonces no es necesario preparar material alguno sobre x. (Allí donde no pica, no vale la pena rascarse.) El mundo nos proporciona un inagotable diluvio de información que bombardea nuestros sentidos, y cuando nos concentramos en toda la información que nos llega, o a la que podemos acceder, a menudo sucumbimos

a la ilusión de que es preciso utilizarla toda, siempre. Pero nuestras capacidades para usar información y nuestros apetitos epistémicos son limitados. Si nuestro cerebro puede satisfacer todos nuestros apetitos epistémicos particulares a medida que éstos surgen, nunca hallaremos un motivo de queja. De hecho, nunca podremos decir que nuestro cerebro nos está proporcionando menos de todo cuanto está a nuestra disposición en el mundo.

Hasta aquí, nos hemos limitado a presentar este principio de economía, aunque no se puede decir que lo hayamos establecido. Como veremos, el cerebro no siempre se aprovecha de esta opción en todos los casos, pero es importante no olvidar esta posibilidad. Todavía no se ha reconocido suficientemente la capacidad de este principio para resolver viejos enigmas.

4. Presentación

En los capítulos que siguen, intentaré dar una explicación de la conciencia. Más precisamente, explicaré los diversos fenómenos que conforman aquello que llamamos conciencia, y demostraré que son todos efectos físicos de las actividades del cerebro; explicaré también cómo han evolucionado estas actividades y de qué manera dan lugar a ilusiones sobre sus poderes y propiedades. Es difícil imaginar cómo puede ser que nuestra mente sea nuestro cerebro, pero no es imposible. A fin de imaginar algo así, es preciso conocer muchos de los descubrimientos científicos sobre el cerebro y, lo que es más importante, es necesario aprender nuevas maneras de pensar. La acumulación de hechos nos ayuda a imaginar nuevas posibilidades, pero los descubrimientos y las teorías de la neurociencia no son suficientes; incluso aquellos que se dedican a la neurociencia se sienten abrumados por los problemas que plantea la conciencia. Con el fin de ampliar su imaginación, junto a los hechos científicos relevantes, he incluido una serie de historias, analogías, experimentos mentales y otras ayudas especialmente diseñadas para proporcionarle nuevas perspectivas, romper con viejos hábitos en la manera de pensar y ayudarle a organizar los hechos en una única visión coherente completamente distinta de la visión tradicional de la conciencia que tendemos a creer. El experimento mental del cerebro en un tarro y la analogía con el juego del psicoanálisis son ejercicios de calentamiento que nos permitirán abordar con más garantías el objetivo principal de este libro: esbozar una teoría de los mecanismos biológicos y *una manera de pensar* sobre estos mecanismos que le pemitirán ver cómo es posible resolver los tradicionales misterios y paradojas de la conciencia.

En la primera parte revisaremos los diversos problemas de la conciencia y estableceremos algunos métodos. Esto es mucho más importante y difícil de lo que uno podría pensar. Muchos de los problemas con que se han encontrado otras teorías son la consecuencia de haber empezado con mal pie, con la intención de hallar una respuesta para las grandes preguntas demasiado pronto. Los novedosos supuestos básicos de mi teoría juegan un pa-

pel fundamental más adelante, lo que nos permitirá posponer la resolución de muchos de los enigmas filosóficos con los que han chocado otros investigadores, hasta que hayamos esbozado una teoría basada en hechos empíricos que presentamos en la segunda parte.

El modelo de las Versiones Múltiples esbozado en la segunda parte es una alternativa al modelo tradicional de la conciencia, que he bautizado con el nombre de Teatro Cartesiano. Requiere una radical revisión del concepto tradicional de «flujo de la conciencia» y, en un primer momento, resulta profundamente antiintuitivo; sin embargo, se va apoderando de uno a medida que ve cómo es capaz de dar cuenta de ciertos problemas sobre el cerebro que hasta hoy habían sido ignorados por los filósofos (y por los científicos). Al considerar con cierto detalle la cuestión de cómo puede haber evolucionado la conciencia, conseguiremos aportar nueva luz sobre determinadas características de nuestra mente que, de otra manera, resultan desconcertantes. En esta segunda parte analizamos también el papel que juega el lenguaje dentro de la conciencia, así como la relación que existe entre el modelo de las Versiones Múltiples con concepciones más tradicionales de la mente humana y con otras concepciones teóricas desarrolladas en el marco del campo multidisciplinar de la ciencia cognitiva. A lo largo de nuestro camino, y hasta que no nos sintamos seguros con los nuevos fundamentos, deberemos resistirnos a la tentadora simplicidad de la visión tradicional.

En la tercera parte, armados de nuevas maneras de guiar nuestra imaginación, podremos afrontar (finalmente) los misterios tradicionales de la conciencia: las extrañas propiedades del «campo fenoménico», la naturaleza de la introspección, las cualidades (o *qualia*) de los estados de la experiencia, la naturaleza del yo o el ego y su relación con los pensamientos y las sensaciones, la conciencia de las criaturas no humanas. Las paradojas que han obstaculizado los debates tradicionales sobre este asunto podrán ahora ser vistas como productos de una *falta de imaginación*, no de «inteligencia», con lo que seremos capaces de eliminar los misterios.

Este libro presenta una teoría que es a la vez empírica y filosófica y, dado que los requisitos que se le imponen son de muy diversa índole, se incluyen dos apéndices que tratan brevemente algunos problemas concretos que surgen tanto desde el punto de vista científico como del filosófico. En el siguiente capítulo, tratamos el problema de cómo podría ser una explicación de la conciencia y de si es lícito querer acabar con todos sus misterios.

PRIMERA PARTE

PROBLEMAS Y MÉTODOS

Capítulo 2

EXPLICAR LA CONCIENCIA

1. La caja de Pandora: ¿es necesario desmitificar la conciencia?

> He aquí unos árboles cuya aspereza conozco, y un agua que saboreo. Estos perfumes de hierba y de estrellas, la noche, ciertos crepúsculos en que el corazón se dilata: ¿cómo negaría yo este mundo cuya potencia y cuyas fuerzas experimento? Sin embargo, toda la ciencia de esta tierra no me dará nada que pueda asegurarme que este mundo es mío.
>
> Me lo describís y me enseñáis a clasificarlo. Me enumeráis sus leyes y en mi sed de saber consiento en que sean ciertas. Desmontáis su mecanismo y mi esperanza aumenta.
>
> (...) ¿Qué necesidad tenía yo de tantos esfuerzos? Las líneas suaves de esas colinas y la mano del crepúsculo sobre este corazón agitado me enseñan mucho más.
>
> ALBERT CAMUS, *El mito de Sísifo*, 1942†

> Dulce es el que trae la naturaleza;
> nuestro intelecto asombrado
> desfigura las formas hermosas de las cosas;
> —asesinamos para ser minuciosos.
>
> WILLIAM WORDSWORTH, «The Tables Turned», 1798*

La conciencia humana es el último de los grandes misterios. Un misterio es un fenómeno para el cual no hemos hallado todavía una manera de pensar. Ha habido muchos otros grandes misterios: el misterio del origen del universo, el misterio de la vida y la reproducción, el misterio del diseño oculto de la naturaleza, los misterios del tiempo, el espacio y la gravedad. Éstas no eran solamente áreas de ignorancia científica, sino motivos para la profunda perplejidad y sorpresa. Todavía no poseemos las respuestas últimas para todos los problemas de la cosmología y la física de partículas, la genética molecular y la teoría de la evolución, pero sabemos cómo pensar sobre ellos. Los misterios no han desaparecido, pero han sido domados. Ya no su-

† Traducción castellana de Luis Echávarri, Alianza, Madrid, 1951 [1988].

* Traducción castellana de Santiago Corugedo y José Luis Chamosa en William Wordsworth y Samuel Coleridge, *Baladas Líricas*, Cátedra, Madrid, 1990.

peran nuestros esfuerzos por comprender los fenómenos, porque ahora sabemos distinguir las malas preguntas de las buenas e, incluso si la aceptación de determinadas respuestas finalmente nos demuestra que estamos completamente equivocados, sabemos cómo seguir buscando nuevas y mejores respuestas.

Con la conciencia, sin embargo, seguimos sumidos en la más profunda de las confusiones. La conciencia se caracteriza por ser el único tema que todavía puede dejar mudos y turbados a los más sofisticados pensadores. Y como ya ocurrió en su momento con los demás misterios, hay muchos que insisten —y esperan— que nunca llegará la desmitificación de la conciencia.

Los misterios son apasionantes, lo que, después de todo, es parte de aquello que hace que la vida sea divertida. A nadie le gusta el gracioso que, saliendo del cine, te cuenta quién es el asesino mientras haces cola para entrar. Una vez alguien ha tirado de la manta, ya nunca volvemos a recuperar ese estado de dulce mistificación que nos embriagaba. Así que, ¡atención lectores!, porque si tengo éxito en mi intento de explicar la conciencia, aquellos que sigan leyendo van a sustituir el misterio por los rudimentos del conocimiento científico de la conciencia, lo cual puede ser un mal negocio para algunos. Dado que algunas personas identifican desmitificación con profanación, imagino que en un principio considerarán este libro como un acto de vandalismo intelectual, como un asalto al último santuario de la humanidad. Me gustaría hacerles cambiar de opinión.

Camus nos da a entender que no necesita de la ciencia, porque puede aprender más de la suave línea de las colinas y de la mano del atardecer, y no soy yo quien va a contradecirle —teniendo en cuenta las preguntas que se hace Camus—. La ciencia no posee todas las buenas preguntas. Tampoco la filosofía. Y, precisamente por este motivo, los fenómenos de la conciencia, que ya son desconcertantes por sí mismos independientemente de las inquietudes de Camus, no tienen por qué ser protegidos de la ciencia, ni de la investigación filosófica desmitificadora en que nos hemos embarcado. A veces la gente, temiendo que la ciencia «asesine para ser minuciosa» como diría Wordsworth, se siente atraída por doctrinas filosóficas que ofrecen algún tipo de garantía en contra de esta invasión. Los recelos que han motivado la aparición de estas doctrinas no carecen de fundamento, independientemente del valor de las mismas. Evidentemente, *podría* ser que la desmitificación de la conciencia constituyera una gran pérdida. Uno de mis objetivos aquí es demostrar que esto *no ocurrirá*: las pérdidas, si las hubiere, se verán compensadas por las ganancias en comprensión —tanto científicas y sociales como teóricas y morales— que una buena teoría de la conciencia nos puede proporcionar.

¿En qué medida *puede* la desmitificación de la conciencia ser algo que debamos lamentar? Puede que resulte ser algo parecido a la pérdida de la inocencia de la edad adulta, que sin lugar a dudas es una pérdida, aunque con evidentes contrapartidas. Consideremos, por ejemplo, lo que ha ocurrido con el amor al hacernos más sofisticados. Somos capaces de compren-

der que un caballero andante quisiera sacrificar su vida por el honor de una princesa con la que apenas había cruzado unas palabras —ésta es una idea que me resultaba particularmente emocionante cuando tenía once o doce años—, pero no es el tipo de estado emocional en el que hoy en día un adulto entra con facilidad. Antaño, la gente poseía un discurso y unas maneras de pensar sobre el amor que actualmente son prácticamente impensables, con la excepción, quizá, de los niños y de los que son capaces de suprimir sus conocimientos de adultos. A todos nos gusta expresar nuestro amor a quienes queremos, y también nos gusta que a su vez ellos lo hagan con nosotros; sin embargo, en tanto que adultos, ya no estamos tan seguros de lo que esto significa como lo estábamos antes, cuando éramos niños y el amor era una cosa sencilla.

¿Saldremos ganando o perdiendo con este cambio de perspectiva? Evidentemente, el cambio no es uniforme. Mientras algunos cándidos adultos siguen elevando la novela rosa a los primeros lugares de la lista de *best-sellers*, nosotros, lectores refinados, consideramos que nos hemos vuelto inmunes a los pretendidos efectos de tales libros: nos hacen reír, y no llorar. Y si a veces nos hacen llorar —como ocurre, muy a pesar nuestro—, nos sentimos turbados por descubrir que aún somos capaces de caer en una trampa tan fácil; pues nos resulta difícil compartir los sentimientos de la heroína que pasa su tiempo preguntándose si realmente ha encontrado el «verdadero amor», como si se tratara de una sustancia diferente (oro emocional, en contraste con el latón o cobre emocionales). Este acto de crecer no afecta solamente a los individuos. También nuestra cultura se ha hecho más refinada o, en todo caso, el refinamiento es un rasgo bastante generalizado dentro de la cultura. A resultas de ello, nuestros conceptos del amor han cambiado, y con ellos se han producido cambios de sensibilidad que actualmente nos impiden experimentar ciertos sentimientos que en su tiempo apasionaron, consternaron o excitaron a nuestros antepasados.

Con la conciencia ha ocurrido algo parecido. Hoy en día hablamos de nuestras decisiones conscientes y de nuestros hábitos inconscientes, de las sensaciones conscientes que experimentamos (a diferencia de los cajeros automáticos, por ejemplo, que no experimentan tales sensaciones), pero ya no estamos tan seguros de qué queremos decir cuando utilizamos tales expresiones. A pesar de que todavía hay pensadores que con gran vehemencia sostienen que la conciencia es algo precioso y genuino (como el amor, como el oro), algo que «esta ahí» y que es muy, muy especial, va ganando cuerpo la sospecha de que todo esto no es más que una ilusión. Quizá los diversos fenómenos que conspiran para crear la sensación de que nos hallamos ante un único y misterioso fenómeno, no tienen más unidad última y esencial que los diferentes fenómenos que contribuyen a la sensación de que el amor es algo simple.

Comparemos el amor y la conciencia con dos fenómenos totalmente distintos: las enfermedades y los terremotos. Nuestros conceptos de enfermedad y terremoto han sufrido cambios notables en los últimos cien años; sin

embargo, las enfermedades y los terremotos son fenómenos en gran medida (pero no completamente) independientes del concepto que nosotros tengamos de ellos. Que hayamos cambiado de opinión sobre las enfermedades no ha contribuido en sí mismo a que las enfermedades desaparecieran o se hicieran menos frecuentes, aunque sí que ha provocado cambios en la medicina y la salud pública que han alterado el modo en que éstas se manifiestan. Es posible que algún día seamos capaces de controlar, o cuando menos predecir, los terremotos, pero su existencia no se verá mayormente afectada por la actitud o la concepción que nosotros tengamos hacia los terremotos. Con el amor es distinto. Ya no es posible que cierta gente refinada «se enamore» de la misma manera en que antes podía uno enamorarse, simplemente porque ya no creen en *esas* maneras de enamorarse. Por ejemplo, yo ya no podría «colarme» por alguien como a los quince años, a menos que sufriera una «regresión hacia la adolescencia» y en el proceso lograra olvidarme o abandonara mucho de lo que creo saber. Afortunadamente para mí, existen otros muchos tipos de amor en los que creer, pero, ¿qué pasaría si no los hubiese? El amor es uno de esos fenómenos que *dependen de sus conceptos*, por decirlo de forma poco precisa por el momento. Existen otros conceptos parecidos, como el de dinero. Si todo el mundo olvidara qué es el dinero, ya no habría dinero; habría fajos de papeletas impresas, discos de metal trabajado, apuntes informatizados de cuentas corrientes, bancos de mármol y granito, pero no habría dinero: ni inflación ni deflación, ni tipos de cambio ni de interés —ni *precios*—. Esa propiedad que poseen aquellos recortes de papel impreso, y que explica —como ninguna otra cosa podría hacerlo— su paso de una mano a otra como consecuencia de diversos intercambios comerciales, desaparecería por completo.

En la concepción de la conciencia que desarrollaré en este libro, veremos que ésta, como el amor y el dinero, es un fenómeno que depende de forma sorprendente de los conceptos que se le han asociado. Aunque, como el amor, posee una compleja base biológica, alguno de sus rasgos más significativos, como el dinero, se transmiten junto con otros valores culturales y no son simplemente inherentes a la estructura física de sus realizaciones particulares. Así pues, si finalmente estoy en lo cierto, y si consigo echar por tierra algunos de estos conceptos, habré conseguido amenazar de extinción a todos los fenómenos de la conciencia que dependen de ellos. ¿Estamos a punto de entrar en un período posconsciente de la conceptualización humana? ¿Debemos temer que algo así ocurra? ¿Es siquiera concebible?

Si el concepto de conciencia acabara por «caer del lado de la ciencia», ¿qué ocurriría con nuestros sentidos de la acción moral y del libre albedrío? Si la experiencia consciente quedara «reducida» a mera materia en movimiento, ¿qué sería de nuestra apreciación del amor, del dolor, de los sueños y de la felicidad? Si los seres humanos conscientes fueran «solamente» objetos materiales animados, ¿cómo podría estar mal o bien cualquier cosa que les hagamos? Éstos son algunos de los temores que alimentan la resis-

tencia y distraen la concentración de aquellos que se enfrentan a un intento de explicar la conciencia.

Estoy convencido de que estos temores son infundados, pero no es evidente que así sea. Hacen subir las apuestas en esta confrontación entre teoría y argumentación que está a punto de comenzar. Existen poderosos argumentos, independientes de estos miedos, que se han esgrimido en contra del tipo de teoría científica materialista que presentaré aquí. No puedo sino reconocer que es a mí a quien corresponde demostrar no sólo que tales argumentos son equivocados, sino también que una amplia aceptación de mi visión de la conciencia no conllevaría en modo alguno consecuencias tan desastrosas. (Y si hubiese descubierto que sí tendría estos efectos, ¿qué habría hecho yo entonces? Pues no habría escrito *este* libro, pero, más allá de esto, ya no lo sé.)

Adoptando una perspectiva más positiva, no debemos olvidar lo que ha ocurrido a raíz de desmitificaciones anteriores. El sentimiento de admiración no disminuye; todo lo contrario, pues se nos aparecen nuevas cosas bellas y visiones aún más deslumbrantes de la complejidad del universo que los protectores del misterio nunca habrían podido concebir. La «magia» de visiones anteriores era, en su mayor parte, el velo que ocultaba flagrantes faltas de imaginación, una aburrida maniobra encerrada en el concepto de un *deus ex machina*. Estos fieros dioses conduciendo carros dorados que cruzaban los cielos son simples personajes de tebeo comparados con la fascinante extrañeza de la cosmología contemporánea, y la complejidad recursiva de los mecanismos de reproducción del ADN convierte el *élan vital* en algo tan interesante como la terrible kriptonita de Supermán. Cuando comprendamos la conciencia —cuando ya no haya misterio—, ésta será diferente, pero seguirá habiendo belleza y más motivos que nunca para el asombro.

2. El misterio de la conciencia

¿Cuál es, pues, el misterio? ¿Qué puede ser más obvio para un individuo que su naturaleza en tanto que sujeto consciente de experiencia, alguien que disfruta de percepciones y sensaciones, que padece el dolor, un pensador consciente, alguien que tiene ideas? Todo esto parece innegable, pero ¿qué objeto del mundo puede ser la conciencia? ¿Cómo es posible que los cuerpos físicos animados que habitan el mundo físico sean capaces de producir tales fenómenos? Éste es el misterio.

El misterio de la conciencia tiene muchas maneras de presentarse. A mí, volvió a golpearme con particular fuerza una mañana, no hace mucho, mientras leía un libro sentado en una mecedora. Aparentemente, había levantado la vista de mi libro, dirigiendo la mirada hacia la ventana, ensimismado, cuando, de repente, la belleza de lo que me rodeaba me distrajo de mis reflexiones teóricas. La luz verde-dorada de los primeros días de primavera penetraba a través de la ventana; las ramas y ramitas del arce del jardín eran

todavía visibles a través de una nube de verdes brotes, formando un elegante dibujo de maravillosa complejidad. La vidriera de la ventana es de cristal antiguo y tiene un pequeño defecto que es apenas perceptible; mientras me balanceaba, esta imperfección provocó una onda de movimientos sincronizados que iba y venía a través del delta de ramas, un movimiento regular que se superponía con gran viveza sobre el desordenado y trémulo reflejo de las ramas movidas por la brisa.

Entonces me di cuenta de que este metrónomo visual en las ramas del árbol seguía acompasadamente las notas del *Concerto grosso* de Vivaldi que había escogido como «música de ambiente» para mi lectura. En un principio pensé que yo mismo había sincronizado inconscientemente mi balanceo con la música —como ocurre cuando uno, sin darse cuenta, acompaña el ritmo con el pie—. Sin embargo, las mecedoras tienen un rango limitado de frecuencias de balanceo, así que, probablemente, la sincronía era sólo una coincidencia, ligeramente depurada por mi preferencia inconsciente por la regularidad, por mantener el paso.

En mi mente, pasé espontáneamente a imaginar los diferentes procesos cerebrales que podrían explicar cómo ajustamos inconscientemente nuestra conducta, incluidos el comportamiento de nuestros ojos y de nuestras facultades de la atención, a fin de «sincronizar» la «banda sonora» con la «imagen». Sin embargo, mis pensamientos se vieron pronto interrumpidos al caer repentinamente en la cuenta de que *lo que yo estaba haciendo* —ese experimentar y pensar simultáneamente sobre lo que estaba experimentando, como he descrito, desde mi posición privilegiada de primera persona— era más difícil de «modelar» que esos procesos inconscientes, entre bastidores, que sin duda se estaban *produciendo dentro de mí* y que, de algún modo, constituían las condiciones causales de lo que estaba haciendo. Era relativamente fácil dar un sentido a esos mecanismos de fondo; lo desconcertante era lo que se producía en un primer plano, los acontecimientos que constituían el centro de atención. Mi pensar consciente y, especialmente, el deleite que me producía esa combinación de luz solar, soleados violines de Vivaldi y ramas susurrantes, además del placer que sentía al pensar en todo ello; ¿cómo es posible que *todo esto* no sea más que algo físico que está ocurriendo en mi cerebro? ¿Cómo es posible que una determinada combinación de sucesos electroquímicos en mi cerebro pudiera sumarse a la delicadeza con que cientos de ramitas se balanceaban al ritmo de la música? ¿Cómo es posible que un acto de procesamiento de la información en mi cerebro *sea* la suave calidez con que yo sentía la luz del sol sobre mi cuerpo? En cuanto a eso, ¿cómo es posible que un evento en mi cerebro sea la imagen mental imperfectamente visualizada por mí de... un acto de procesamiento de la información en mi cerebro? Parece realmente imposible.

Parece como si los acontecimientos que *son* mis pensamientos conscientes y mis experiencias no pudieran ser acontecimientos cerebrales, sino que deberían ser *otra cosa*; sin duda algo causado o producido por mis acontecimientos cerebrales, pero algo que está ahí además de esos acontecimientos,

algo hecho de una sustancia diferente y localizado en un espacio diferente. Bueno, ¿y por qué no?

3. Los atractivos de la sustancia mental

Veamos qué ocurre cuando seguimos por este camino que sin duda resulta tan atractivo. En primer lugar, quisiera que usted realizara un pequeño experimento. Deberá cerrar los ojos, imaginar algo y, después, una vez se haya usted formado su imagen mental y la haya verificado cuidadosamente, deberá responder a las preguntas que aparecen más abajo. No lea las preguntas hasta que no haya seguido la siguiente instrucción: cierre los ojos e imagine, lo más detalladamente posible, una vaca de color violeta.

¿Ya? Entonces:

1. ¿La vaca miraba hacia la derecha, hacia la izquierda o estaba de frente?
2. ¿Estaba rumiando?
3. ¿Eran visibles sus ubres?
4. ¿El color, era violeta claro o violeta oscuro?

Si siguió las instrucciones, entonces podrá responder a todas las preguntas sin necesidad de reconstruir algún detalle retrospectivamente. Si, por el contrario, encuentra que las preguntas le resultan difíciles de responder, entonces no se tomó usted muy en serio lo de imaginarse una vaca de color violeta. Quizá se limitó a pensar, perezosamente, «me estoy imaginando una vaca de color violeta» o «pongamos que *esto* es imaginarse una vaca de color violeta», o algo por el estilo.

Ahora, realicemos un segundo ejercicio: cierre los ojos e imagine, lo más detalladamente posible, una *vaca de color amarillo.*

En esta ocasión es probable que usted sea capaz de responder a las tres primeras preguntas sin demasiadas dudas, y quizá pueda afirmar con cierta certeza qué tono de amarillo —amarillo pastel, amarillo canario, ocre— cubre los flancos de su vaca imaginaria. Sin embargo, ahora quisiera considerar una pregunta diferente:

5. ¿Qué diferencia hay entre imaginar una vaca de color violeta e imaginar una vaca de color amarillo?

La diferencia es obvia: la primera vaca imaginaria es de color violeta y la segunda es de color amarillo. Puede que haya otras diferencias, pero la esencial es ésta. El problema es que, dado que estas vacas son sólo vacas imaginarias y no vacas reales, ni retratos de vacas pintados sobre una tela, ni formas con apariencia de vaca en una pantalla de televisión, es difícil hacerse una idea sobre qué es de color violeta en el primer caso y qué de color amarillo en el segundo. No hay nada con forma de vaca en su cerebro (o en

su globo ocular) que se vuelva violeta en un caso y amarillo en el segundo; incluso si así fuera, no sería de mucha ayuda, porque dentro de su cráneo está muy oscuro y, además, usted no tiene ojos para mirar ahí dentro y ver los colores.

Hay eventos que tienen lugar en su cerebro y que están íntimamente ligados a sus actos particulares de imaginación. No queda descartado, pues, que en un futuro próximo un investigador del cerebro, al examinar los procesos que se produjeron en su cerebro como respuesta a mis instrucciones, fuera capaz de descifrarlos a fin de poder comprobar la veracidad de sus respuestas a las preguntas 1-4:

«¿Miraba la vaca hacia la izquierda? Eso parece. El patrón de excitación neuronal correspondiente a la cabeza de la vaca es consistente con una presentación del cuadrante visual superior izquierdo; observamos, además, señales de detección del movimiento oscilatorio de un hertzio, lo que indica actividad rumiante; no pudimos detectar, sin embargo, actividad en los grupos de representación del complejo-ubre, y, una vez calibrados los potenciales suscitados con los perfiles de detección del color del sujeto, podemos aventurar que éste ha mentido sobre el color de la vaca; casi con toda certeza, creemos poder afirmar que la vaca imaginaria era de color marrón».

Supongamos que todo esto fuera cierto; supongamos también que hubiese llegado la era de la lectura científica del pensamiento. Sin embargo, no parece probable que el misterio quedase resuelto de este modo: ¿qué es lo que es de color marrón cuando imaginamos una vaca de color marrón? No lo es el evento en el cerebro que los científicos han calibrado con su acto de experimentar el marrón. El tipo y la localización de las neuronas que intervienen, sus conexiones con otras partes del cerebro, la frecuencia o la amplitud de actividad, los neurotransmisores químicos liberados —ninguna de estas propiedades es la propiedad misma de la vaca «en su imaginación»—. Y al imaginar usted una vaca (usted no miente, los científicos nos lo confirman), en ese preciso instante una vaca imaginaria empieza a existir. Algo, en algún lugar, debe haber tenido esas propiedades en ese momento. Se ha reproducido una vaca, no en el medio de la sustancia cerebral, sino en el medio de... la sustancia mental. ¿Dónde, si no?

La sustancia mental debe ser, pues, «aquello de lo que están hechos los sueños», y, aparentemente, posee algunas propiedades sorprendentes. Ya hemos mencionado una de ellas de forma tangencial, aunque muestra una notable resistencia a ser definida. Por el momento, diremos que la sustancia mental siempre *tiene un testigo*. Como hemos señalado, el problema de los eventos cerebrales es que, independientemente de lo próximos que éstos estén a los eventos de nuestro flujo de conciencia, siempre tienen una desventaja insalvable: *nunca hay nadie que pueda presenciarlos*. Los eventos que se producen en su cerebro, al igual que los eventos que tienen lugar en su estómago o en su hígado, no suelen producirse ante los ojos de nadie, ni tam-

poco existe una diferencia en la manera que tienen de producirse cuando hay un observador y cuando no lo hay. Los eventos de la conciencia, por otra parte, son «por definición» presenciados; son *experimentados* por un *experimentador*, y es eso precisamente lo que hace que sean lo que son: *eventos conscientes*. Un evento experimentado no es algo que pueda tener lugar de forma aislada; debe ser la experiencia de *alguien*. Para que se produzca un pensamiento, alguien (alguna mente) debe pensarlo; para que se produzca un dolor, alguien debe sentirlo; y para que una vaca de color violeta empiece a existir «en la imaginación», alguien debe imaginarla.

Por el contrario, el problema con los cerebros es, parece, que cuando miramos en su interior, vemos que *ahí no hay nadie*. Ninguna parte del cerebro es el pensador que piensa o el sentidor que siente; tampoco todo el cerebro al completo parece ser un mejor candidato para cumplir ese papel tan especial. Estamos ante un asunto delicado. ¿Piensan los cerebros? ¿Ven los ojos? ¿O, quizá las personas ven con sus ojos y piensan con sus cerebros? ¿Hay alguna diferencia? ¿Es éste un mero problema «gramatical» o nos revela una de las principales fuentes de confusión? La idea de que hay un yo (o una persona o, para el caso, un alma) distintos del cerebro o el cuerpo está profundamente arraigada en nuestra manera de hablar y, por tanto, en nuestra manera de pensar.

«Yo tengo un cerebro.»

Parece ser algo incontestable. Sin embargo, no parece ser sinónimo de

«Este cuerpo tiene un cerebro» (y un corazón, y dos pulmones, etc.).

o de

«Este cerebro se tiene a sí mismo.»

Es bastante natural pensar en «el yo y su cerebro» (Popper y Eccles, 1977) como dos cosas diferentes, con propiedades diferentes, independientemente de cuánto dependan el uno del otro. Si el yo es distinto del cerebro, entonces debe de estar hecho de sustancia mental. En latín, una cosa pensante es una *res cogitans*, término que Descartes hizo popular cuando ofreció lo que él consideraba ser una prueba irrefutable de que él, en tanto que objeto manifiestamente pensante, no podía ser su cerebro. Lo que sigue es una de las versiones de este argumento de Descartes, y es realmente convincente:

> Examiné después atentamente lo que yo era, y viendo que podía fingir que no tenía cuerpo alguno y que no había mundo ni lugar alguno en el que yo me encontrase, pero que no podía fingir por ello que no fuese, sino al contrario, por lo mismo que pensaba en dudar de la verdad de las otras cosas, se seguía muy cierta y evidentemente que yo era, mientras que, con sólo dejar

> de pensar, aunque todo lo demás que había imaginado fuese verdad, no tenía ya razón alguna para creer que yo era, conocí por ello que yo era una sustancia cuya esencia y naturaleza toda es pensar, y que no necesita, para ser, de lugar alguno, ni depende de cosa alguna material [...]. (*Discurso del método*, 1637, IV.)*

Así pues, hemos descubierto dos tipos de cosas que querríamos elaborar a partir de la sustancia mental: la vaca de color violeta que no está en el cerebro, y aquello que efectivamente realiza la acción de pensar. Pero existen otros poderes especiales que quisiéramos atribuir a la sustancia mental.

Supongamos que un vinatero decidiera sustituir a sus catadores por una máquina. Un «sistema experto» informatizado para el control de calidad y la clasificación de los vinos es algo que prácticamente está dentro de las posibilidades de la tecnología actual. Sabemos lo suficiente sobre la química del vino como para construir los transductores que harían las veces de las papilas gustativas y de los receptores olfativos del epitelio (los que proporcionan la «materia prima» —los estímulos de entrada— de los sentidos del gusto y el olfato). No se conoce con precisión de qué manera se combinan estos estímulos de entrada para producir nuestras experiencias, pero se están realizando progresos notables. La investigación en el campo de la visión ha avanzado bastante más. Los trabajos sobre visión del color sugieren que imitar la idiosincrasia, delicadeza y fiabilidad humanas en el componente de detección del color de la máquina constituiría un desafío tecnológico de enormes proporciones, pero no es algo imposible. Así pues, podemos imaginar que es posible utilizar la elaborada salida de esos transductores sensoriales y sus mecanismos de comparación como entrada de complejas rutinas de clasificación, descripción y evaluación. Vierta una muestra de vino en el embudo y, en unos minutos o en unas horas, el sistema mandará imprimir un ensayo químico junto con un comentario: «Un Pinot chispeante y aterciopelado, aunque algo falto de cuerpo», o algo por el estilo. Una máquina como ésta incluso podría superar a los mejores catadores en precisión y consistencia, pero es seguro que por muy «sensible» y «preciso» que fuera el sistema, parece que nunca sentirá ni disfrutará lo que nosotros cuando bebemos vino.

¿Es esto tan evidente? Según las diversas ideologías agrupadas bajo la etiqueta de *funcionalismo*, si somos capaces de reproducir la completa «estructura funcional» del sistema cognitivo de un catador de vinos humano (memoria, objetivos, íntimas animadversiones, etc., incluidos), seremos entonces capaces de reproducir *todas* las propiedades mentales, incluidos el disfrute y el deleite que hacen de saborear un buen vino algo que muchos de nosotros *apreciamos*. En principio no hay diferencia, nos dice el funcionalista, entre un sistema hecho de moléculas orgánicas y otro hecho de sili-

* Traducción castellana de Manuel García Morente, Espasa Calpe Mexicana (Selecciones Austral), México, 1976, pág. 62.

cio, mientras ambos *hagan el mismo trabajo.* Los corazones artificiales no tienen por qué estar hechos de tejido orgánico, ni tampoco tienen por qué estarlo los cerebros artificiales (al menos en principio). Si todas las funciones de control del cerebro de un catador de vinos pudieran ser reproducidas en un *chip* de silicio, veríamos *ipso facto* también reproducido el deleite que éste siente.

Es muy posible que finalmente triunfe alguna versión de funcionalismo (de hecho, en este libro defenderemos una de esas versiones del funcionalismo), aunque en una primera impresión nos pueda parecer una barbaridad. No parece posible que una simple máquina, por muy bien que reproduzca los procesos cerebrales del catador de vinos, sea capaz de apreciar el vino, o una sonata de Beethoven, o un partido de baloncesto. Para poder apreciar se necesita la conciencia, algo que una simple máquina no tiene. Sin embargo, es evidente que el cerebro *es* una especie de máquina, un órgano que, como el corazón o los pulmones, en última instancia es susceptible de una explicación mecánica de sus capacidades. Ello puede hacer que nos parezca atractivo pensar que no es el cerebro quien posee la capacidad de apreciar; *eso* es responsabilidad (o privilegio) de la mente. Reproducir los mecanismos cerebrales en una máquina basada en el silicio no reproduciría la capacidad real de apreciar, sino tan sólo una ilusión o un simulacro de esa capacidad.

Así pues, la mente consciente no sólo es el lugar donde están los colores y los olores que percibimos, ni tampoco es solamente la cosa pensante. Es el lugar donde se lleva a cabo la apreciación. Es el árbitro último que decide por qué algo es importante. Quizá todo ello sea consecuencia del hecho de que se suele asumir que la mente consciente es el origen de nuestras acciones intencionales. Es razonable suponer —¿no?— que si *hacer cosas que importan* depende de la conciencia, atribuir *importancia* a algo (disfrutar, apreciar, sufrir, preocuparse) debe también depender de la conciencia. Si un sonámbulo «inconscientemente» causa algún daño, no es responsable de ello porque hasta cierto punto sería lícito defender que él no lo hizo. Los movimientos de su cuerpo están íntimamente imbricados con las cadenas causales que tienen como consecuencia ese daño, pero no constituyen una acción por su parte, del mismo modo que tampoco sería así si hubiera causado el daño por el mero hecho de caerse de la cama. La simple complicidad corporal no es suficiente para constituir una acción intencional, ni tampoco lo es la complicidad corporal *bajo el control de las estructuras cerebrales,* ya que el cuerpo del sonámbulo se halla bajo el control manifiesto de las estructuras cerebrales del propio sonámbulo. Lo que nos falta añadir es la conciencia, el ingrediente especial que convierte los meros *acontecimientos* en *actos.*[1]

No podemos culpar al Vesubio si mata a nuestro ser amado en una erupción. En un caso como éste ni el resentimiento (Strawson, 1962) ni el desdén

1. Véase mi libro *La libertad de acción* (1984a), capítulo 4, para una amplia discusión de este asunto.

son actitudes aceptables, a menos que lleguemos a convencernos de que el Vesubio, contrariamente a la concepción más extendida actualmente, es un agente consciente. Sin duda resulta extrañamente reconfortante para nuestro dolor el librarnos a actitudes de este tipo, maldecir la «furia» del huracán, renegar contra el cáncer que injustamente trunca la vida de un niño, o abominar de «los dioses». Originalmente, decir que algo era «animado» en oposición a «inanimado» equivalía a decir que poseía una alma (*anima* en latín). Quizás el pensar que las cosas que tanto nos afectan son animadas tenga otra finalidad, además del simple consuelo; es posible que sea una estrategia enraizada en nuestro diseño biológico, un atajo para ayudar a nuestros cerebros abrumados por la prisa a organizarse y a pensar en aquello en que debemos pensar si queremos sobrevivir.

Es probable que poseamos una tendencia innata a tratar toda cosa cambiante como si ésta tuviera una alma (Stafford, 1983; Humphrey, 1983b, 1986), sin embargo, por muy natural que sea esta actitud, ahora sabemos que atribuir una alma (consciente) al Vesubio es ir demasiado lejos. El lugar preciso por donde hay que trazar la línea divisoria es un problema complejo sobre el que volveremos más adelante, pero para nosotros, parece que la conciencia es precisamente aquello que nos diferencia de los meros «autómatas». Los «reflejos» corporales son «automáticos» y mecánicos; es probable que requieran la intervención de circuitos en el cerebro, pero no la intervención de la mente consciente. Resulta bastante natural pensar en nuestros cuerpos como si fueran polichinelas que «nosotros» controlamos «desde el interior». Hago que el polichinela salude con la mano a la audiencia moviendo mi dedo; hago que mi dedo se mueva... ¿cómo? ¿Moviendo mi alma? Esta idea plantea problemas bien conocidos, lo cual no impide que hasta cierto punto parezca correcta: a no ser que detrás de todo acto haya una mente consciente, no existe un *auténtico* agente que lo lleve a cabo. Cuando pensamos en nuestras mentes de esta manera, es como si descubriéramos a nuestro «yo interior», nuestro «yo real». Este yo real no es nuestro cerebro; es lo que *posee* a nuestro cerebro («el yo y *su* cerebro»). Sobre el escritorio de Harry Truman, en el Despacho Oval de la Casa Blanca, había un cartel que rezaba: «Cargadme a mí con el muerto».* Sin embargo, no parece haber una parte del cerebro *a la que cargar con el muerto*, el origen último de la responsabilidad moral a la cabeza de una cadena de mandos.

Resumiendo, pues, hemos encontrado cuatro motivos por los que creer en la sustancia mental. No parece que la mente consciente pueda ser simplemente el cerebro, ni alguna de sus partes, porque nada en el cerebro podría

1. ser el medio en que se reproduce la vaca de color violeta;
2. ser la cosa pensante, el yo del «pienso, luego existo»;
3. apreciar el vino, odiar el racismo, amar a alguien, ser capaz de atribuir importancia a algo;
4. actuar con responsabilidad moral.

* Literalmente, «The buck stops here». [N. del T.]

Toda teoría aceptable de la conciencia humana debe ser capaz de dar cuenta de estos cuatro argumentos en favor de la existencia de una sustancia mental.

4. Por qué el dualismo es un proyecto estéril

La idea de que la mente no es lo mismo que el cerebro, de que no está compuesta de la materia ordinaria sino de un tipo especial de sustancia, es el dualismo, una concepción que, a pesar de los persuasivos argumentos que acabamos de exponer, goza de una merecida mala reputación hoy en día. Desde el clásico ataque de Gilbert Ryle (1949) a lo que éste denominó el «dogma del espíritu en la máquina» de Descartes, los dualistas se han puesto a la defensiva.[2] La concepción dominante, expresada y defendida de muy diversas maneras, es el *materialismo*: es decir, la idea de que sólo hay un tipo de sustancia, la *materia*, la sustancia física de la química, la física y la fisio-

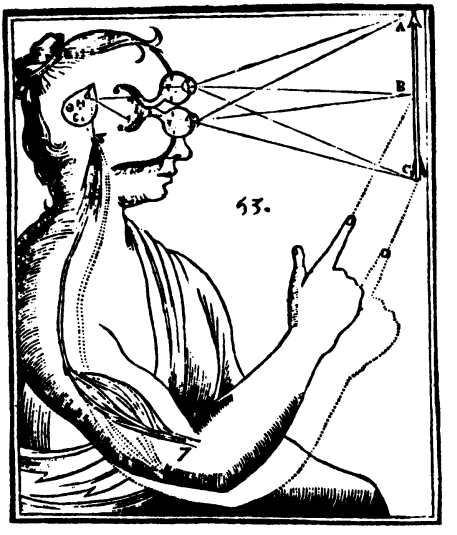


Figura 2.1

2. Algunas almas valerosas (y no cabe duda de que ninguno de ellos rechazaría tal calificativo) han opuesto alguna resistencia: tanto el trabajo de Arthur Koestler, con el desafiante título de *The Ghost in the Machine* (1967), como el de Popper y Eccles *The Self and Its Brain* (1977) han sido escritos por autores de una eminencia incuestionable; otras dos defensas, iconoclastas y penetrantes, del dualismo son los trabajos de Zeno Vendler *Res Cogitans* (1972) y *The Matter of Minds* (1984).

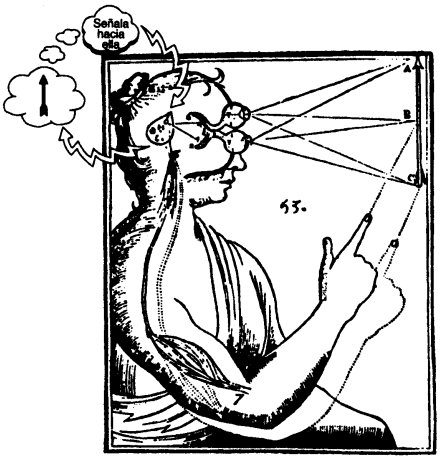


Figura 2.2

logía, y según la cual la mente no es nada más que un fenómeno físico. En pocas palabras, la mente es el cerebro. Según los materialistas, podemos explicar (en principio) cualquier fenómeno mental con los mismos principios, leyes y materias primas físicas que nos sirven para explicar la radioactividad, la deriva continental, la fotosíntesis, la reproducción, la nutrición y el crecimiento. Uno de los más duros cometidos de este libro será el de explicar la conciencia sin ceder nunca al canto de sirena del dualismo. Pero, ¿qué tiene de tan malo el dualismo? ¿Por qué ha caído en desgracia?

La principal objeción al dualismo ya era bien conocida por Descartes en el siglo XVII, y es justo decir que ni él ni ningún otro dualista han sido nunca capaces de vencerla. Si la mente y el cuerpo son sustancias distintas, deben de todos modos ser capaces de interrelacionarse; los órganos sensoriales del cuerpo, a través del cerebro, deben *informar* a la mente, deben enviarle o presentarle percepciones, ideas o datos de algún tipo, y la mente, a su vez, después de reflexionar sobre ello, debe *dirigir* al cuerpo en las acciones apropiadas (incluida el habla). Por este motivo, es corriente referirse a esta concepción con el nombre de interaccionismo cartesiano o dualismo interaccionista. En la formulación de Descartes, el locus de interacción en el cerebro era la glándula pineal o epífisis, la cual aparece representada en el diagrama del propio Descartes, con una medida un tanto exagerada, como un óvalo en el centro de la cabeza.

A fin de ilustrar con más claridad el problema del interaccionismo, podemos sobreimponer un esbozo del resto de la teoría de Descartes sobre su propio diagrama (figura 2.2).

La percepción consciente de la flecha sólo se produce cuando el cerebro

ha transmitido su mensaje a la mente, y el dedo de la persona sólo puede señalar a la flecha una vez que la mente ha transmitido sus órdenes al cuerpo. ¿Cómo se transmite exactamente la información de la glándula pineal a la mente? Como no tenemos ni la más remota idea (todavía) de qué propiedades tiene la sustancia mental, tampoco estamos capacitados (todavía) para averiguar de qué manera puede ésta verse afectada por los procesos físicos que emanan del cerebro. Por el momento, pues, nos permitiremos ignorar esas señales dirigidas hacia arriba, y nos concentraremos en las señales de retorno, las instrucciones de la mente hacia el cerebro. Éstas, *ex hypothesi*, no son físicas; no son ondas de luz o de sonido, ni rayos cósmicos ni flujos de partículas subatómicas. No tienen asociada ninguna energía física ni una masa. ¿De qué manera, pues, consiguen intervenir sobre lo que ocurre en las células cerebrales a las que tienen que afectar, si la mente debe tener alguna influencia sobre el cuerpo? Uno de los principios fundamentales de la física es que cualquier cambio de trayectoria que sufra una entidad física es una aceleración que requiere un gasto de energía, pero ¿de dónde proviene esta energía? Es este principio de conservación de la energía lo que explica la imposibilidad física de las «máquinas de movimiento perpetuo», y es precisamente este mismo principio el que aparentemente viola el dualismo. Este conflicto entre la física estándar y el dualismo, que viene produciéndose repetidamente desde los tiempos de Descartes, es contemplado como el más fatal e ineludible de los defectos del dualismo.

Como sería de esperar, se han explorado y expuesto numerosas e ingeniosas exenciones técnicas basadas en lecturas bastante sofisticadas de la física relevante para la cuestión, aunque sin ganar muchos conversos a la causa. De hecho, el desconcierto de los dualistas tiene una causa mucho más simple de lo que sugiere la alusión a unas presuntas leyes de la física. Se trata de la misma incoherencia que los niños observan —pero toleran con su jovial fantasía— en las historietas de Casper el fantasma bueno (figura 2.3, pág. 48). ¿Cómo es posible que Casper pueda pasar a través de las paredes y *al mismo tiempo* recoger una sábana que se ha caído? ¿Cómo puede eludir la sustancia mental toda medición física y *al mismo tiempo* controlar el cuerpo? Un espíritu en la máquina no nos será de mucha ayuda en nuestra teoría a menos que sea un espíritu capaz de mover objetos, como un fantasma ruidoso capaz de volcar una lámpara o dar un portazo. Sin embargo, cualquier cosa que pueda mover un objeto físico es a su vez un objeto físico (quizá un algo físico extraño y poco estudiado, pero físico al fin).

¿Qué podemos decir entonces de la alternativa de suponer que la sustancia mental es un tipo especial de materia? En la época victoriana, durante las sesiones de espiritismo, los médium a menudo hacian surgir de la nada una cosa que denominaban «ectoplasma», una extraña sustancia viscosa que supuestamente constituía la materia básica del mundo de los espíritus. Una sustancia, no obstante, que podía guardarse en un tarro de cristal, escurridiza, húmeda y capaz de reflejar la luz igual que la materia ordinaria. Estos pequeños fraudes no deben, por ello, ser un obstáculo para preguntarnos

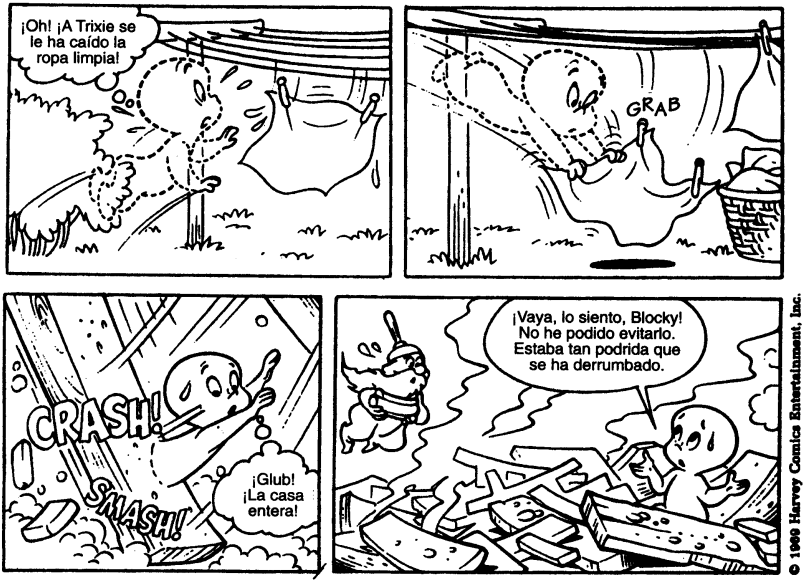


Figura 2.3

seriamente si la sustancia mental podría ser algo que está más allá de los átomos y las moléculas que forman nuestro cerebro, sin dejar de ser, por ello, un tipo de materia susceptible de ser investigada de modo científico. La ontología de una teoría es el catálogo de cosas y tipos de cosas que esa teoría supone que existen. La ontología de las ciencias físicas incluyó en su momento el «calórico» (la sustancia de la que estaba hecho el calor) y el «éter» (la sustancia que llenaba el espacio y medio por el que se transmitían las ondas de luz del mismo modo que el aire y el agua son medios por los que se transmiten las ondas de sonido). Hoy en día nadie cree seriamente en la existencia de tales cosas, mientras que es normal conceder a los neutrinos, la antimateria y los agujeros negros un lugar dentro de la ontología de las ciencias. Quizá sea necesario ampliar la ontología de las ciencias físicas a fin de explicar los fenómenos de la conciencia.

En esta línea de pensamiento, el físico y matemático Roger Penrose ha propuesto una revolución parecida dentro de la física en su libro *La nueva mente del emperador* (1989). Aunque dudo que haya conseguido tener éxito en su intento de iniciar una revolución,[3] creo que merece la pena señalar que Penrose se ha cuidado mucho de caer en la trampa del dualismo. ¿Cuál es la diferencia? Penrose deja claro que su objetivo al proponer una revolu-

3. Véase «Murmurs in the Cathedral» (Dennett, 1989c), mi reseña de su libro.

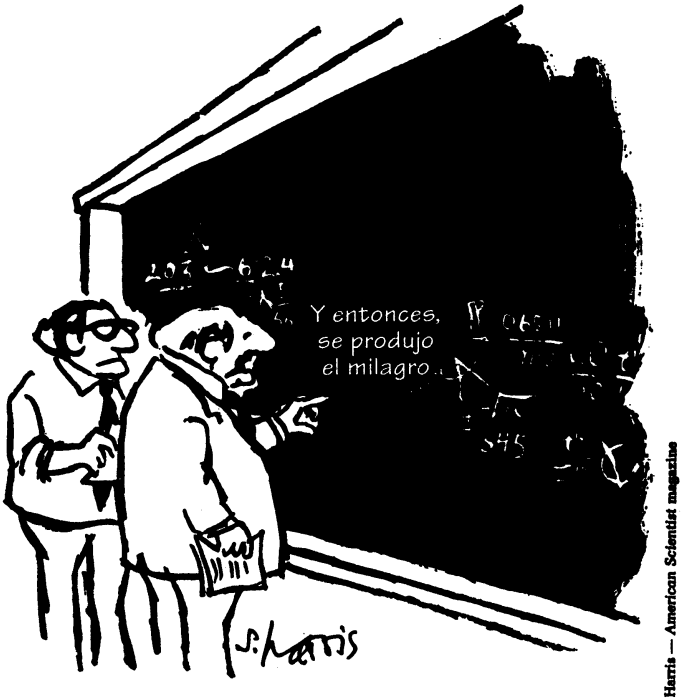
ción es hacer que la mente consciente sea *más* accesible a la investigación científica, y no menos. Con toda seguridad, no es accidental que los pocos dualistas que abiertamente reconocen su postura hayan admitido, con gran candor y naturalidad, que carecen por completo de una teoría sobre cómo funciona la mente; es algo, insisten, que está por encima de la capacidad de comprensión de los humanos.[4] Permanece, así, la sospecha oculta de que la característica más atractiva de la sustancia mental es su promesa de seguir siendo tan misteriosa como para mantener en jaque a la ciencia para siempre.

Esta actitud fundamentalmente anticientífica es, a mi modo de ver, la causa principal que hace perder al dualismo toda su legitimidad, y es también el motivo por el cual en este libro adoptaré la postura dogmática de que el dualismo debe evitarse *a toda costa*. No es que yo piense que soy capaz de ofrecer una prueba definitiva de que el dualismo, en todas sus formas, es falso o incoherente, sino, simplemente, que, atendiendo a la forma en que el dualismo se refugia en el misterio, considero que *aceptar el dualismo equivale a darse por vencido* (como en la figura 2.4, pág. 50).

Existe un amplio consenso sobre este punto, pero el consenso es tan amplio como frágil, y no hace más que disimular algunas grietas preocupantes en el muro del materialismo. Científicos y filósofos pueden haber alcanzado un cierto grado de consenso en favor del materialismo, pero como veremos, desembarazarse de las viejas visiones dualistas es mucho más difícil de lo que creen los materialistas contemporáneos. La tarea de hallar sustitutos aceptables para las tradicionales imágenes dualistas exigirá llevar a cabo un cierto número de ajustes radicales en nuestras maneras de pensar, ajustes que en un principio resultarán tan contraintuitivos para los científicos como para los profanos.

No considero que sea mala señal que mi teoría aparezca en un principio como algo claramente reñido con la tradición. Al contrario, no debemos esperar de una buena teoría de la conciencia que nos resulte fácil y cómoda de aceptar, que sea una de esas teorías que inmediatamente «hace que se nos encienda la bombilla» y nos hace exclamar, con orgullo contenido: «¡Claro! ¡Ya lo sabía yo! ¡Una vez te lo demuestran es evidente!». Si realmente existiera la posibilidad de desarrollar una teoría como ésta, ya lo habríamos hecho hace tiempo. Los misterios de la mente han permanecido vivos durante tanto tiempo, y nuestros progresos han sido tan pocos, que existen muchas posibilidades de que mucho de lo que todos estábamos de acuerdo en considerar como obvio quizá no lo sea tanto. En breve presentaré mis propios candidatos a este puesto.

4. Eccles ha propuesto que la mente no física está compuesta por millones de «psicones», que interactúan con millones de «dendrones» (regiones de las células piramidales) en el córtex; cada psicón se corresponde aproximadamente con lo que Descartes o Hume llamarían una idea —como la idea de rojo, la idea de redondo o la de caliente—, pero aparte de esta descomposición mínima, Eccles no tiene nada que añadir sobre las partes, actividades, principios de acción u otras propiedades de la mente no física.



Creo que debería ser usted un poco más explícito aquí, en el segundo paso.

© 1975 Sidney Harris — American Scientist magazine

Figura 2.4

Hoy en día, algunos investigadores del cerebro, quizá una gran mayoría, siguen afirmando que, para ellos, el cerebro no es más que otro órgano, como los riñones o el páncreas, que debe ser descrito y explicado única y exclusivamente con las seguras herramientas que nos ofrecen las ciencias físicas y biológicas. Nunca se les ocurriría mencionar la mente ni lo «mental» en el curso de sus tareas profesionales. Para otros investigadores con inclinaciones más teóricas existe un nuevo objeto de estudio, la mente/cerebro (Churchland, 1986). Este popular término de nuevo cuño expresa claramente el materialismo dominante entre estos investigadores, quienes no sienten pudor alguno de admitir ante el mundo y ante sí mismos que lo que hace que el cerebro sea tan fascinante es que de una manera u otra es la mente. Pero incluso entre estos investigadores permanece una cierta reticencia a enfrentarse con los grandes problemas, un deseo de posponer hasta un futuro indeterminado las embarazosas preguntas que plantea la naturaleza de la conciencia.

Aun cuando esta actitud es plenamente razonable, un modesto reconocimiento al valor de la estrategia del divide y vencerás tiene el efecto negativo de distorsionar la imagen de algunos nuevos conceptos que han surgido dentro de lo que hoy se denomina *ciencia cognitiva*. Casi todos los investigadores en este campo, tanto si se sienten afines a la neurociencia como a la psicología o la inteligencia artificial, tienden a posponer los problemas relacionados con la conciencia al restringir su atención a sistemas «periféricos» y «subordinados» de la mente/cerebro. Estos sistemas, se supone que sirven y alimentan a un oscuro e imaginario «centro» en el que se producen el «pensamiento consciente» y la «experiencia». Esto suele tener la consecuencia de dejar que gran parte del trabajo de la mente se lleve a cabo «en el centro», lo cual, a su vez, hace que se subestime la «cantidad de comprensión» que debe producirse en los sistemas relativamente periféricos del cerebro (Dennett, 1984b).

Por ejemplo, muchos teóricos tienden a pensar en los sistemas perceptivos como proveedores de información de «entrada» para un ruedo central del pensamiento, el cual a su vez proporciona «control» o «dirección» para aquellos sistemas relativamente periféricos encargados del movimiento del cuerpo. Se supone también que este ruedo central saca provecho del material almacenado en los diversos sistemas subsidiarios de la memoria. Esta idea de que existen divisiones teóricas importantes entre presuntos subsistemas tales como la «memoria a largo plazo» y el «razonamiento» (o «planificación») es, no obstante, más un producto de la estrategia del divide y vencerás que algo que se pueda encontrar en la naturaleza. Como veremos enseguida, esta atención exclusiva a subsistemas específicos de la mente/cerebro a menudo causa una especie de miopía teórica que impide a los investigadores ver que sus modelos aún presuponen que, en algún lugar, oculto en el oscuro «centro» de la mente/cerebro, hay un Teatro Cartesiano, un lugar al que «todo va a parar» y donde se produce la conciencia. Puede que ésta sea una buena idea, una idea inevitable, y hasta que veamos, con detalle, por qué no lo es, el Teatro Cartesiano seguirá atrayendo la atención de un sinnúmero de teóricos deslumbrados por una ilusión.

5. El reto

En el apartado anterior he observado que si el dualismo es lo máximo a que podemos aspirar, entonces nunca podremos comprender la conciencia humana. Algunas personas están convencidas de que así será siempre, hagamos lo que hagamos. Este derrotismo, hoy, en un momento en que podemos sacar provecho de infinidad de avances científicos, me parece ridículo, incluso patético, aunque puede que sea la triste realidad. Quizá sea verdad que la conciencia no pueda explicarse, pero ¿cómo vamos a saberlo si no lo intenta alguien? Creo que ya comprendemos muchas de las piezas del rompecabezas —de hecho, la mayoría de ellas—, y sólo es necesario hacer-

las encajar con un poco de ayuda por mi parte. Aquellos que quieren defender a la mente de la ciencia deberían desearme suerte, ya que si son ellos los que están en lo cierto, entonces mi proyecto está condenado al fracaso, pero si hago bien mi trabajo, mi derrota servirá para dilucidar por qué la ciencia siempre será insuficiente. Así, tendrán en sus manos finalmente su argumento definitivo en contra de la ciencia, y yo les habré hecho todo el trabajo sucio.

Las normas básicas de mi proyecto son sencillas:

1. *Nada de tejidos milagrosos.* Intentaré explicar cualquier rasgo enigmático de la conciencia humana dentro del marco de la ciencia física contemporánea; en ningún momento recurriré a fuerzas, sustancias o poderes orgánicos inexplicables o desconocidos. En otras palabras, pretendo ver hasta dónde se puede llegar ciñéndome a los límites conservadores de la ciencia estándar, reservando como último recurso el apelar a una revolución dentro del materialismo.
2. *Nada de anestesias fingidas.* Se ha dicho de los conductistas que fingen estar anestesiados, que fingen ser ajenos a las experiencias que sabemos que comparten con nosotros. Si quiero negar la existencia de algún rasgo controvertido de la conciencia, es tarea mía el *demostrar* que se trata de una ilusión.
3. *Nada de regatear con los detalles empíricos.* Intentaré dar cuenta de todos los hechos científicos correctamente, en la medida que son conocidos hoy en día, aunque exista un cierto desacuerdo sobre qué avances resistirán el paso del tiempo. Si me ciñera exclusivamente a «aquellos hechos que ya aparecen en los manuales», no podría sacar partido de los descubrimientos recientes más reveladores (si eso es lo que realmente son). Y si la historia reciente ha de servirnos de ejemplo, no dejaré por ello de fomentar inconscientemente algunas falsedades. Algunos de los «descubrimientos» sobre la visión por los cuales David Hubel y Torstein Wiesel merecieron el Premio Nobel en 1981 han empezado a aclararse sólo recientemente; y la famosa teoría «retinex» de la visión del color de Edwin Land, que durante más de veinte años ha sido considerada como un hecho establecido por la mayoría de los filósofos de la mente y otros especialistas, cada vez goza de menos predicamento entre los investigadores de la visión.[5]

Así pues, dado que como filósofo mi cometido es el de determinar las *posibilidades* (y refutar cualquier presunta imposibilidad), me limitaré a dibujar esbozos de teorías en lugar de desarrollar teorías completas y verificadas empíricamente. El esbozo de una teoría o un modelo sobre cómo *podría* hacer algo el cerebro, puede convertir una perplejidad en un proyecto

5. El filósofo C. L. Hardin, en un apéndice de su libro *Color for Philosophers: Unweaving the Rainbow* (1988), lleva a cabo una excelente revisión sobre la vigencia de la teoría de Land.

de investigación: si este modelo no funciona, ¿funcionará entonces alguna variación más realista del mismo? (La explicación del fenómeno de las alucinaciones que esbozamos en el capítulo 1 sería un ejemplo de ello.) Tales esbozos son susceptibles de verse directa y explícitamente refutados por medios empíricos; sin embargo, si usted pretende sostener que mi esbozo no es una explicación *posible* de un determinado fenómeno, tendrá que demostrar qué aspectos del mismo *deben* ser abandonados y qué cosas no *puede* hacer el mencionado esbozo; si, por el contrario, usted se limita a afirmar que ciertos detalles de mi modelo pueden no ser correctos, no podré más que estar de acuerdo. Por ejemplo, lo malo del dualismo cartesiano no es que Descartes eligiera la glándula pineal —en vez del tálamo o la amígdala— como locus de interacción con la mente, sino *la idea* misma de que existe un punto de interacción entre la mente y el cerebro. Está claro que lo que debe ser considerado como regateo cambia a medida que la ciencia avanza y diferentes teóricos establecen diferentes criterios. Intentaré, pues, pecar de exceso de detalle, no sólo con el fin de recalcar el contraste con la filosofía de la mente tradicional, sino también para proporcionar a los críticos que adoptan un punto de vista empírico una buena diana sobre la que disparar.

En este capítulo hemos presentado las características básicas del misterio de la conciencia. Su característica principal es precisamente ese misterio, un rasgo vital, quizá, sin el cual no puede sobrevivir. Dado que ésta es una posibilidad que goza de cierto predicamento, la prudencia tiende a favorecer aquellas doctrinas que ni siquiera se plantean explicar la conciencia, ya que la conciencia es algo muy importante para nosotros. El dualismo, la idea de que un cerebro no puede ser una cosa pensante y, por tanto, de que una cosa pensante no puede ser un cerebro, resulta tentador por diversos motivos, pero debemos resistirnos a esa tentación; «adoptar» el dualismo significa simplemente aceptar la derrota sin admitirlo. La adopción del materialismo no resuelve por sí misma todos los enigmas de la conciencia, los cuales, por otra parte, tampoco se siguen de simples inferencias extraídas de las ciencias del cerebro. En cierto modo, el cerebro tiene que ser la mente, pero a menos que podamos llegar a ver con cierto detalle cómo es esto posible, nuestro materialismo no explicará la conciencia, se limitará solamente a prometer que, un buen día, lo hará. Como he sugerido, tales promesas no podrán cumplirse a menos que aprendamos a abandonar la mayor parte del legado de Descartes. Al mismo tiempo, independientemente de lo que nuestras teorías materialistas lleguen a explicar, no explicarán la *conciencia* si desatendemos aquellos hechos sobre la experiencia que tan íntimamente conocemos «desde el interior». El próximo capítulo lo dedicaremos a establecer un inventario de estos hechos.

Capítulo 3

UNA VISITA AL JARDÍN FENOMENOLÓGICO

1. Bienvenidos al fenome

Imaginemos que un loco se presentara ante nosotros y tratara de convencernos de que los animales no existen. Podríamos intentar sacarlo de su error llevándolo al zoo y diciéndole «¡Mira! ¿Qué son esas cosas que ves si no animales?». Con ello no podríamos esperar llegar a curarlo, pero como mínimo tendríamos la satisfacción de establecer ante nuestros ojos la magnitud de su locura. Supongamos, sin embargo, que su reacción fuese respondernos «Ya, yo ya sé que estas cosas existen —los leones, los avestruces, las boas constríctor—, pero ¿qué os hace pensar que lo que vosotros llamáis animales son *animales*? La verdad es que no son más que *robots* cubiertos de piel —bueno, algunos también están cubiertos de plumas y otros de escamas». Puede que esto siga siendo una locura, pero es un tipo de locura distinto y más defendible. Resulta que nuestro loco tiene una idea revolucionaria acerca de la naturaleza última de los animales.[1]

Los zoólogos son expertos en la naturaleza última de los animales, y los jardines zoológicos —zoos para abreviar— sirven al propósito educativo de poner al alcance de la gente los objetos de su conocimiento. Si los zoólogos llegaran a descubrir que nuestro loco estaba en lo cierto (en cierto sentido), un zoo les sería muy útil en sus intentos de explicar su descubrimiento. Dirían, «Resulta que los animales —ya saben, esas cosas que vemos en los zoos— no son lo que creíamos que eran. De hecho, son tan diferentes que no deberíamos llamarlos animales. Así que, resumiendo, los animales, en el sentido tradicional del término, no *existen*».

Los filósofos y los psicólogos utilizan con frecuencia el término *fenomenología* como genérico que engloba todos aquellos elementos —la fauna y

1. El propio Descartes tenía una concepción parecida sobre los animales. Sostenía que éstos no eran más que complejas máquinas. También los cuerpos humanos, e incluso los cerebros humanos, no eran más que máquinas. Es sólo gracias a nuestras mentes no mecánicas y no físicas que los seres humanos (y sólo los seres humanos) son inteligentes y conscientes. Era ésta una idea bastante sutil, la mayor parte de la cual sería adoptada sin problemas por muchos zoólogos hoy en día, pero era demasiado revolucionaria para los contemporáneos de Descartes, quienes, como pueden ustedes imaginarse, se dedicaron a caricaturizarla y después se mofaron de la caricatura. Siglos después, estas calumnias contra Descartes siguen siendo propagadas alegremente por los que encuentran inconcebible —o al menos intolerable— la idea de una explicación mecanicista de la conciencia. Para un análisis bastante esclarecedor, véase Leiber (1988).

la flora, podríamos decir— que habitan el mundo de nuestra experiencia consciente: pensamientos, olores, picores, dolores, vacas imaginarias de color violeta, intuiciones y todo lo demás. Este uso del término tiene orígenes ligeramente distintos que merece la pena recordar. En el siglo XVIII, Kant distinguía entre «fenómenos», las cosas tal como nos aparecen, y «*noumenos*», las cosas como son en sí mismas. Con el desarrollo de las ciencias naturales o físicas en el siglo XIX, el término *fenomenología* pasó a designar simplemente todo estudio descriptivo de cualquier materia, de forma neutral o preteórica. La fenomenología del magnetismo, por ejemplo, ya había sido iniciada por William Gilbert en el siglo XVI, pero su explicación tuvo que esperar a los descubrimientos sobre la relación entre el magnetismo y la electricidad llevados a cabo en el siglo XIX, y al trabajo teórico de Faraday, Maxwell y otros. En alusión a esta dicotomía entre observación precisa y explicación teórica, la escuela o movimiento filosófico conocido como Fenomenología (con F mayúscula) nació a principios del siglo XX en torno a la figura de Edmund Husserl. Su objetivo era establecer unas nuevas bases para la filosofía (y, de hecho, para todo el conocimiento) a partir de una técnica especial de introspección. De acuerdo con esta técnica, el mundo exterior y todas sus implicaciones y presuposiciones deben ser puestas «entre paréntesis» en un acto particular de nuestra mente al que se denominó *epojé*. El resultado de este proceso era un estado investigativo de la mente gracias al cual se suponía que el fenomenólogo podía acceder a los objetos puros de la experiencia consciente, denominados *noemas*, sin verse influido por las distorsiones y prejuicios fruto de la teoría y la práctica. Igual que ha ocurrido con otros intentos de eliminar la interpretación y someter así los hechos básicos de la conciencia a una observación rigurosa, como por ejemplo el movimiento impresionista en las artes y las psicologías introspeccionistas de Wundt, Titchener y otros, la Fenomenología ha sido incapaz de hallar un único método con el que todo el mundo estuviera de acuerdo.

Así, mientras podemos decir que hay zoólogos, no podemos decir realmente que haya fenomenólogos: expertos en la naturaleza de las cosas, por todos reconocidos, que nadan en la corriente de la conciencia. Podemos, no obstante, seguir la práctica habitual reciente de adoptar el término (con f minúscula) como genérico para designar todos aquellos elementos de la experiencia consciente que deben ser explicados.

Hace algún tiempo, publiqué un artículo titulado «Sobre la ausencia de fenomenología» (1979a) en el que intentaba argumentar en favor del segundo tipo de locura: los elementos de los que está compuesta la conciencia son tan diferentes de lo que se había creído hasta ahora, que ya no deberíamos utilizar los términos tradicionales para referirnos a ellos. Pero resultó ser una sugerencia tan escandalosa para muchas personas («¡Cómo es posible que estemos equivocados sobre la naturaleza de nuestra vida interior!»), que intentaron rechazarla como un caso del primer tipo de locura («¡Dennett cree que los dolores, los aromas y los ensueños no existen!»). Es evidente que esto

no es más que una caricatura de mi concepción, pero es también una actitud que resulta tentador adoptar. Mi problema fue el de no tener a mano ningún jardín fenomenológico —un fenome, para abreviar— para utilizarlo en mis explicaciones. Lo que yo quise decir es, «Resulta que las cosas que nadan en la corriente de la conciencia —ya saben: los dolores, los aromas, los ensueños, las imágenes mentales y los arrebatos de ira y lujuria, los típicos moradores de un fenome— no son lo que creíamos que eran. La verdad es que son tan diferentes, que les deberíamos buscar nuevos nombres».

Hagamos, pues, una breve visita al jardín fenomenológico, sólo para estar seguros de que sabemos de qué estamos hablando (aunque no sepamos aún cuál es la naturaleza última de las cosas). Por fuerza, no podrá ser más que una visita superficial e introductoria; nos limitaremos a identificar los diversos moradores del jardín, a decir algunas palabras sobre ellos y a plantear algunas preguntas, antes de encomendarnos a tareas teóricas más serias que abordaremos en el resto del libro. Dado que pronto voy a presentar desafíos radicales en contra del pensamiento tradicional, no quisiera que nadie pensara que desconozco por completo las cosas maravillosas que habitan en la mente de los *demás*.

Nuestro fenome se divide en tres partes: (1) *experiencias del mundo «exterior»*, tales como imágenes, sonidos, olores, sensaciones resbaladizas y rasposas, sensaciones de frío y calor, y sensaciones sobre la posición de los miembros de nuestro cuerpo; (2) *experiencias del mundo puramente «interno»*, tales como imágenes fantasiosas, las visiones y sonidos interiores fruto de nuestros sueños y nuestras conversaciones con nosotros mismos, recuerdos, buenas ideas y corazonadas repentinas; (3) *experiencias emotivas o «de afecto»* (por utilizar un término un tanto torpe del que gustan mucho los psicólogos), entre las que encontramos, por un lado, los dolores corporales, las cosquillas, las «sensaciones» de hambre y sed, pero también arrebatos emocionales de rabia, felicidad, odio, vergüenza, lujuria, asombro, un amplio abanico que va desde las visitaciones menos corpóreas del orgullo, la ansiedad, el remordimiento, el distanciamiento irónico, el arrepentimiento, el pánico o la frialdad, pasando por una zona intermedia de rabia, felicidad, odio, vergüenza, lujuria o asombro.

No voy a comprometerme con la validez de esta clasificación tripartita entre lo externo, lo interno y lo afectivo. Como una vieja casa de fieras que pone los murciélagos con los pájaros y los delfines con los peces, esta taxonomía se basa más en la semejanza superficial y en una tradición dudosa que en una supuesta íntima relación entre los distintos fenómenos. Por algún sitio tenemos que empezar, no obstante, y cualquier taxonomía que nos proporcione una cierta orientación es útil para evitar que pasemos por alto alguna especie.

2. Nuestra experiencia del mundo exterior

Empecemos por nuestros sentidos más elementales: el gusto y el olfato. Es bien sabido que nuestras papilas gustativas son sensibles sólo a lo dulce, a lo ácido, a lo salado y a lo amargo, y que en gran medida «saboreamos con la nariz». Éste es el motivo por el cual los alimentos pierden su sabor cuando estamos resfriados. El epitelio nasal es al sentido del olfato lo que la retina es a la vista. Las células epiteliales son de muchos tipos, y cada una es sensible a un tipo distinto de las muchas moléculas que flotan en el aire. Lo que importa es la *forma* de estas moléculas, que flotan hasta la nariz y, como si de pequeños interruptores se tratara, accionan una u otra célula del epitelio. Con frecuencia es posible detectar las moléculas incluso en concentraciones muy bajas del orden de unas pocas partes por cada mil millones. Otros animales tienen mucho mejor olfato que nosotros, y no sólo son capaces de discriminar muchos más olores en concentraciones aún más bajas (el sabueso es un ejemplo típico de ello), sino que también poseen una capacidad mejor desarrollada para introducir las dimensiones de espacio y tiempo en la detección de olores. Nosotros podemos llegar a sentir la presencia de moléculas de formaldehído en una habitación, pero si lo hacemos, no olemos que esas moléculas se disponen en una estela que discurre en línea recta o se concentran en una región del espacio con moléculas perceptibles individualmente flotando por la habitación; toda la habitación o, en todo caso, un rincón de esa habitación, estará inundada por el olor. No debe sorprendernos que esto sea así: las moléculas *flotan a la deriva* y pasan, más o menos al azar, por nuestros orificios nasales, de modo que su llegada a determinados puntos del epitelio proporciona información muy escasa acerca de su procedencia; no ocurre lo mismo con los fotones, que fluyen en líneas rectas ópticas a través del pequeño orificio del iris hasta llegar a una determinada dirección retinal, la cual se proyecta geométricamente sobre una fuente exterior o una trayectoria-fuente. Si la resolución de nuestra visión fuese tan pobre como la de nuestro olfato, un pájaro volando sobre nuestras cabezas haría que el cielo adoptara un condición *pajariforme* por unos instantes. (Algunas especies tienen una visión así de pobre: es decir, su capacidad de resolución y de discriminación no va más allá del hipotético caso que hemos descrito. Otra cosa es qué efecto produce en el animal el tener una visión tan pobre; volveremos sobre este asunto en un capítulo posterior.)

Nuestros sentidos del gusto y del olfato están fenomenológicamente acoplados, como también lo están nuestros sentidos del tacto y de la cinestesia, el sentido de la posición y el movimiento de nuestras extremidades y otras partes del cuerpo. «Sentimos» las cosas al tocarlas, agarrarlas o ejerciendo algún tipo de presión sobre ellas, pero las sensaciones conscientes resultantes, pese a que intuitivamente parecen ser «traducciones» directas de la estimulación de los receptores táctiles bajo nuestra piel, también son el producto de un elaborado proceso de integración de datos procedentes de muy distintas fuentes. Tápese los ojos y coja una varilla (un bolígrafo o un lápiz).

Toque algunas cosas a su alrededor con la vara y notará que puede distinguir las diferentes texturas de los objetos sin demasiado esfuerzo, como si su sistema nervioso tuviera sensores en la punta de la vara. Se precisa un pequeño esfuerzo para atender al modo en que se siente el palito entre nuestros dedos, a cómo vibra o a cómo vence el rozamiento al entrar en contacto con las diferentes superficies. Este intercambio entre la vara y los receptores táctiles (ayudado en muchos casos por sonidos casi imperceptibles) proporciona la información que el cerebro integra en un reconocimiento consciente de la textura del papel, el cartón, la lana o el cristal. Sin embargo, estos complejos procesos de integración distan mucho de ser transparentes a la conciencia; es decir, no nos damos cuenta —no podemos darnos cuenta— de cómo lo hacemos. Pongamos un caso aún más extremo, piense en cómo percibe usted que el firme de la carretera está resbaladizo por la presencia de una mancha de aceite bajo las ruedas de su coche en el momento de tomar una curva. El punto focal fenomenológico de contacto es aquel punto en que el neumático toca con el suelo, y no un punto del cuerpo de usted, sentado y vestido, o un punto en el asiento del coche o en sus manos enguantadas y agarradas al volante.

Ahora, sin destaparse los ojos, deje la varilla y haga que alguien le haga tocar objetos de porcelana, plástico, madera pulida y metal. Todos ellos son lisos y suaves al tacto, pero usted no encontrará ninguna dificultad en distinguirlos (y no porque usted posea receptores especializados para la porcelana o el plástico en las manos). La diferencia de conductividad del calor parece ser uno de los factores más importantes, pero no es esencial: le resultará bastante sorprendente la facilidad con que podrá usted distinguir las diferentes superficies «sintiendo» sólo con la varilla. Esta posibilidad debe depender de vibraciones en la varilla o de diferencias indescriptibles —pero detectables— en los crujidos y chasquidos que se oyen. Sin embargo, *es* como si nuestras terminaciones nerviosas estuvieran en la varilla, porque *percibimos* las diferentes superficies *en* la punta de la varilla.

Consideremos ahora el oído, cuya fenomenología consta de todos aquellos sonidos que podemos percibir: música, palabras habladas, golpes/estallidos, silbidos, sirenas, gorjeos y chasquidos. Los teóricos que se ocupan del oído a menudo se sienten tentados de «poner a tocar a la pequeña orquesta en nuestra cabeza». Esto es un error, y para estar seguros de que lo identificaremos y lo evitaremos, quisiera ilustrarlo con la ayuda de un pequeño cuento.

Había una vez, a mediados del siglo XIX, un inventor de ojos enloquecidos que se enzarzó en una discusión con Filo, un filósofo testarudo. El inventor tenía como objetivo construir una máquina capaz de «grabar» automáticamente y después «reproducir» con alta «fidelidad» una orquesta y unos coros interpretando la Novena sinfonía de Beethoven. «Vaya tontería», dijo Filo. «Eso es imposible. Puedo llegar a imaginarme un ingenio mecánico que grabe la secuencia de golpes de tecla en un piano y que luego controle la reproducción de esa secuencia en un piano especialmente preparado para

ello —podría hacerse con un rollo de papel perforado, por ejemplo—, pero piense usted en la enorme variedad de sonidos y modos de producir estos sonidos que incluye una interpretación de la Novena de Beethoven. Hay cientos de voces humanas de diferentes tonos y timbres, docenas de instrumentos de cuerda, de viento —de metal y de madera—, de percusión. El dispositivo capaz de reproducir tal variedad de sonidos juntos sería una monstruosidad tan difícil de manejar que superaría al más potente de los órganos de iglesia; y si tuviera que hacerlo con la "alta fidelidad" que usted sugiere, no cabe duda de que, literalmente, debería incluir un equipo de esclavos humanos para interpretar las partes vocales, y lo que usted llama la "grabación" de una determinada interpretación, con todos sus matices, debería tener cientos de partituras —una por cada músico— con miles, quizá millones, de anotaciones.»

La argumentación de Filo es, aún hoy, bastante convincente; es sorprendente que todos esos sonidos puedan ser superpuestos fielmente gracias a una serie transformada de Fourier en una sola línea ondulada, grabada sobre un disco de vinilo o representada magnéticamente sobre una cinta u ópticamente sobre la banda sonora de una película. Es aún más sorprendente que un mero cono de papel, al que hace oscilar un electroimán dirigido por la misma línea ondulada, pueda hacer justicia tanto al sonar de una trompeta como al rasgar de las cuerdas de un banjo, al habla humana o al sonido de una botella de vino llena al estrellarse contra el suelo. Filo no pudo imaginarse algo tan poderoso, y tomó su falta de imaginación por una intuición de lo que es necesario.

La «magia» de la serie transformada de Fourier nos abre todo un abanico de nuevas posibilidades en las que pensar, aunque merece la pena señalar que no elimina por sí misma el problema que tanto preocupaba a Filo; simplemente pospone su resolución. Pues mientras nosotros, tan refinados, nos reímos de Filo ante su incapacidad por comprender de qué manera puede grabarse y reproducirse el patrón de compresión y rarificación del aire que estimula nuestro oído, nuestras sonrisas condescendientes desaparecerán de nuestra cara tan pronto como nos planteemos la *siguiente* pregunta: ¿Qué ocurre con la señal una vez que el oído la ha recibido?

Desde el oído, un aluvión de señales moduladas viaja en reata (aunque ahora más o menos analizadas y descompuestas en corrientes paralelas, y recordando siniestramente a los centenares de partituras de Filo) hacia el interior, hacia el oscuro centro del cerebro. Como en el caso de la línea ondulada grabada en el vinilo del disco, no podemos decir que estas cadenas de señales sean *sonidos oídos*: son secuencias de impulsos electroquímicos viajando por los axones de las neuronas. ¿No debería acaso existir algún punto central en el cerebro donde estas señales controlan las acciones del poderoso órgano del teatro de la mente? ¿Dónde, si no, se *traducen* finalmente estas señales sin tonalidad en sonidos oídos subjetivamente?

No queremos buscar en nuestro cerebro lugares que vibren como cuerdas de guitarra, como tampoco queremos hallar puntos en el cerebro que

se vuelvan violetas cuando pensamos en una vaca violeta. Así sólo nos encontraríamos ante un callejón sin salida, ante lo que Gilbert Ryle (1949) llamaría un error categorial. Pero ¿qué *podemos* encontrar entonces en el cerebro que nos permita sentir la satisfacción de haber llegado al final de la cuestión del problema de la experiencia auditiva?[2] ¿Cómo es posible que un conjunto de propiedades físicas de los eventos que se producen en el cerebro constituya —o simplemente explique— las fantásticas propiedades de los sonidos que oímos?

En un primer momento, tales propiedades no parecen susceptibles de ser analizadas o, por utilizar un término favorito de los fenomenólogos, parecen ser *inefables*. Sin embargo, algunas de estas propiedades aparentemente atómicas y homogéneas pueden convertirse en entidades compuestas y susceptibles de ser descritas. Tome una guitarra y pulse la cuerda de mi bajo o superior sin presionar en ningún traste. Escuche el sonido detenidamente. ¿Posee componentes descriptibles o es un sonido único e inefablemente guitarrístico? Muchos optarían por la segunda manera de describir su fenomenología. Pulse ahora la cuerda y, cuidadosamente, deslice un dedo sobre el traste de octava a fin de crear un «armónico» alto. De repente, escuchará usted un nuevo sonido: «más puro» y, evidentemente, una octava más alto. Algunos insisten en afirmar que éste es un sonido totalmente nuevo, mientras otros describen la experiencia diciendo que «ha caído la parte baja de la nota» —quedando sólo la parte superior—. Pulse la cuerda abierta por tercera vez. En esta ocasión podrá usted escuchar, con una claridad sorprendente, el tono armónico que había sido aislado la segunda vez que pulsó la cuerda. La homogeneidad y la inefabilidad de la primera experiencia ha desaparecido, sustituida por una dualidad directamente aprehensible y claramente descriptible como la de cualquier acorde.

La diferencia entre ambas experiencias es sorprendente; pero la complejidad aprehendida en el tercer caso estaba *ahí* desde el principio (percibida y discriminada). Los estudios en este ámbito han demostrado que es sólo gracias al complejo patrón de armónicos que es usted capaz de reconocer el sonido como el de una guitarra y no, por ejemplo, como el de un laúd o el de un arpa. Estas investigaciones pueden sernos de ayuda para *dar cuenta* de las diversas propiedades de las experiencias auditivas, analizando las componentes informativas y los procesos que las integran, y permitiéndonos predecir e incluso sintetizar experiencias auditivas determinadas. Sin embargo, estos trabajos siguen sin abordar la cuestion de a qué *equivalen* tales propiedades. ¿Por qué suena *así* el patrón de tonos armónicos de una guitarra y *asá* el patrón de tonos de un laúd? Aún no hemos respondido a esta pregunta, sin bien hemos podido suavizarla al demostrar que al menos cier-

2. Esta pregunta retórica implica, para algunos, una única y rotunda respuesta: ¡nada! Por ejemplo, McGinn (1989) defiende su derrotista respuesta considerando una serie de opciones posibles, aunque consigue ignorar precisamente aquellas posibilidades que desarrollaremos en capítulos posteriores.

tas propiedades inefables permiten, después de todo, un mínimo de trabajo analítico y descriptivo.[3]

El estudio de los procesos de la percepción auditiva parece indicar que existen mecanismos especializados para descifrar diferentes tipos de sonidos, algo así como los componentes imaginarios de la máquina reproductora de sonidos de Filo. Los sonidos del habla, por ejemplo, parecen ser tratados por lo que un ingeniero denominaría un mecanismo especializado. La fenomenología de la percepción del habla sugiere que en el cerebro se produce una reestructuración total del estímulo de entrada, a través de un mecanismo muy parecido a la mesa de mezclas de un ingeniero de sonido, en la que los distintos canales se mezclan, se potencian y se ajustan para crear un «master» en estéreo del cual se sacarán las copias grabadas en distintos medios.

Por ejemplo, percibimos el habla en nuestra lengua materna como una secuencia de palabras separadas por breves silencios. Es decir, poseemos un claro sentido de las fronteras entre palabras, las cuales no vienen marcadas por líneas coloreadas ni se indican con pitidos o chasquidos. Así pues, ¿qué puede ser una frontera entre palabras si no un silencio de duración variable —como los silencios que separan las letras en el código Morse? Cuando los investigadores piden a los sujetos de un experimento que detecten y evalúen los espacios entre palabras, éstos no tienen ninguna dificultad en hacerlo. Efectivamente, parece que tales espacios existen. Sin embargo, si analizamos el perfil de energía acústica de la señal de entrada, las regiones de baja energía (los espacios de tiempo más próximos al silencio) no siempre se corresponden con las fronteras entre palabras. La segmentación de los sonidos del habla es un proceso que impone fronteras en función de la estructura gramatical del lenguaje, no en función de la estructura física de la onda acústica (Liberman y Studdert-Kennedy, 1977). Ello permite explicar por qué percibimos el habla en lenguas que no conocemos como un torrente de sonidos revueltos y sin segmentar: los mecanismos especializados

3. ¿Por qué un la por debajo del do medio y un la por encima del mismo do (una octava más alto) suenan igual? ¿Qué es lo que hace que ambos sean las? ¿Qué propiedad inefable, el timbre de un la, comparten? Cuando dos tonos están separados por una octava (y, por tanto, nos suenan «igual, pero diferente»), la frecuencia fundamental de uno es exactamente el doble de la frecuencia fundamental del otro. La frecuencia típica de un la por debajo del do es de 220 ciclos por segundo; la de un la una octava más alto (el «la de concierto») es de 440 ciclos por segundo. Cuando ambas notas suenan a la vez, las notas que están separadas de ellas por una octava o más estarán en fase. ¿Será ésta la explicación del misterio del inefable parentesco de estos sonidos? «De ningún modo. ¿Por qué motivo deberían sonar así unas notas en fase?» Las notas que no están en fase no suenan así, pero suenan iguales desde otros puntos de vista (el del timbre, por ejemplo), cuya explicación es distinta en función de las relaciones existentes entre las frecuencias de vibración que producen. Una vez hemos descrito las diversas maneras en que dos notas pueden sonar igual y diferente, las hemos correlacionado con sus propiedades físicas y con los efectos que producen en nuestro aparato auditivo, podemos incluso predecir, con cierta precisión, cómo nos sonarán nuevas notas (por ejemplo, las producidas por un sintetizador electrónico). Si todo esto no explica ese parentesco inefable, ¿qué nos queda por explicar? (Abordaremos con detalle este asunto tan conocido en el capítulo 12.)

del «estudio de sonido» del cerebro carecen de la información gramatical necesaria para detectar los segmentos apropiados, así que lo máximo que pueden hacer es devolver una versión, apenas alterada, de la señal de entrada.

Cuando percibimos el habla, somos conscientes de algo más que de la identidad y la categoría gramatical de las palabras. (Si así fuera, seríamos incapaces de distinguir entre si estamos oyendo o leyendo las palabras.) Las palabras están claramente demarcadas, ordenadas e identificadas, pero también están revestidas de diversas propiedades sensibles. Por ejemplo, acabo de oír la voz claramente británica de mi amigo Nick Humphrey, levemente desafiante, sin llegar a ser burlona. Parece que puedo *oír* su sonrisa; en mi experiencia permanece la sensación de que la risa se escondía detrás de esas palabras, pronta a emerger como el sol que se esconde detrás de las nubes. Las propiedades de las que somos conscientes no sólo son las subidas y bajadas de la entonación, sino también asperezas, silbidos y balbuceos, por no hablar de los tonos afilado de la displicencia, trémulo del miedo o plano de la depresión. Y como pudimos observar en el caso de la guitarra, lo que en un principio parecen propiedades atómicas y completamente homogéneas, a menudo, con un poco de experimentación, pueden ser aisladas y analizadas. Sin apenas esfuerzo, reconocemos la entonación interrogativa de una pregunta —y distinguimos una entonación interrogativa británica de una entonación interrogativa americana—, pero se requiere una cierta experiencia sobre las variaciones de un tema para llegar a describir con seguridad y precisión las diferentes curvas melódicas que dan lugar a los diversos sabores auditivos.

La del «sabor» parece ser aquí la metáfora más apropiada; sin duda, porque nuestra capacidad para analizar sabores es muy limitada. Las conocidas pero aún sorprendentes demostraciones de que gustamos con nuestra nariz no hacen sino mostrar que nuestros sentidos del gusto y el olfato son tan toscos, que incluso tenemos dificultades para identificar la ruta a través de la cual recibimos la información. Tal evidencia no se restringe al caso del gusto y el olfato: nuestra capacidad para oír tonos de muy baja frecuencia, como por ejemplo esas notas bajas y profundas de los órganos de iglesia, aparentemente se debe más a la capacidad de sentir las vibraciones en nuestro cuerpo que a captarlas con nuestros oídos. Resulta sorprendente descubrir que una determinada «calidad de fa sostenido, exactamente dos octavas por debajo del fa sostenido más bajo que yo puedo cantar» puede *oírse* con la parte baja de mi espalda y no con mis orejas.

Para terminar, pasemos a ocuparnos brevemente de la vista. Cuando tenemos los ojos abiertos, recibimos la impresión de hallarnos ante un amplio campo —a menudo denominado campo fenoménico o campo visual— dentro del cual aparecen las cosas, en color y a distintas profundidades o distancias de nosotros, en movimiento o en reposo. Con cierta ingenuidad, vemos los diversos rasgos que experimentamos como propiedades objetivas de las cosas exteriores, observadas «directamente» por nosotros; no obstante, ya de niños, pronto reconocemos una categoría intermedia de elementos

—brillos, destellos, reflejos, siluetas borrosas— que sabemos que son producto de la interacción entre los objetos, la luz y nuestro aparato visual. Seguimos, sin embargo, viendo estos elementos como algo que está «ahí afuera» y no en nosotros, con algunas excepciones: el dolor que sentimos al mirar directamente al sol o a una luz muy brillante cuando nuestros ojos están adaptados a la oscuridad, o ese girar vertiginoso del campo fenoménico cuando estamos mareados. Tales sensaciones difícilmente pueden ser descritas de otra manera que como «sensaciones en los ojos»; se asemejan más a las presiones y los picores que sentimos cuando nos frotamos los ojos que a las propiedades normales de las cosas exteriores que vemos.

De entre las muchas cosas que se pueden ver ahí en el mundo físico están las imágenes pictóricas. Las imágenes son un caso tan claro de cosa-creada-para-ser-vista, que tendemos a olvidar que son una incorporación reciente a nuestro entorno visible, sólo unas pocas decenas de miles de años. Gracias al arte y al artificio humanos, hoy en día vivimos rodeados de cuadros, mapas y diagramas, en movimiento y en reposo. Estas imágenes físicas, que son sólo uno de los muchos tipos de «materia prima» para los procesos de percepción visual, se han convertido inevitablemente en el modelo de lo que es el «producto final» de la percepción visual: «imágenes en la cabeza». Tenemos una fuerte inclinación a decir, «pues, claro que el resultado de la *visión* es una imagen en la cabeza (o en la mente). ¿Qué podría ser si no? En cualquier caso, ¡seguro que no es una melodía o un sabor!». Trataremos esta curiosa aunque omnipresente enfermedad de la imaginación de muchas maneras antes de haber acabado, pero empezaremos con una advertencia: las pinacotecas para ciegos no sirven de mucho, y por el mismo motivo, las imágenes en la cabeza necesitarán de ojos para ser apreciadas (por no hablar de una buena iluminación). Supongamos, pues, que existen unos ojos de la mente para percibir las imágenes en la cabeza. ¿Qué decir entonces de las imágenes en la cabeza de la cabeza producidas por esos ojos internos? ¿Cómo evitar una regresión infinita de imágenes y de espectadores? Sólo podemos romper esa regresión si descubrimos un espectador cuya percepción evite la creación de una nueva imagen que a su vez necesite de un espectador. ¿No será que el punto en que debemos romper la regresión es precisamente en el primer paso?

Afortunadamente, existen motivos independientes para mostrarse escéptico ante esta concepción de la visión basada en imágenes en la cabeza. Si en la visión intervinieran imágenes en la cabeza con las que estamos (nuestros yos interiores) íntimamente relacionados, ¿acaso dibujar no sería entonces más fácil? Piensen en lo difícil que es hacer un dibujo realista de, pongamos por caso, una rosa en un jarrón. Ahí esta la rosa, grande como la vida, a unos pocos metros, enfrente suyo, supongamos que un poco hacia la izquierda de su cuaderno de dibujo. (De verdad quiero que se imaginen ustedes esto con cuidado.) Todos los detalles visibles de la rosa real son claros, precisos e íntimamente accesibles para usted o, por lo menos, eso parece, porque el aparentemente tan simple proceso de encajar una copia de to-

dos estos detalles, en blanco y negro y dos dimensiones, en una hoja de papel situada unos cuantos grados hacia la derecha se convierte en un reto de tal envergadura que la mayoría de la gente pronto abandona y decide que no es capaz de dibujar. La traducción de tres dimensiones a dos es algo que resulta particularmente difícil a la gente, lo cual no deja de ser sorprendente, ya que lo que en un principio parece ser la traducción inversa —es decir, ver una imagen bidimensional realista *como si fuera* una situación u objeto tridimensional— se lleva a cabo involuntariamente y sin esfuerzo. De hecho, es la misma dificultad que tenemos para suprimir esa interpretación inversa lo que convierte incluso el proceso de copiar un dibujo lineal en una tarea de una cierta dificultad.

No se trata solamente de «coordinación entre mano y ojo», ya que los que saben bordar o montar relojes de bolsillo con gran destreza pueden mostrarse como auténticos ineptos en el momento de copiar un dibujo. Podría decirse que es más un problema de coordinación entre ojo y *cerebro*. Los que dominan la técnica saben que ésta requiere el dominio de ciertos hábitos de atención, de trucos como el de desenfocar ligeramente los ojos a fin de poder suprimir la contribución de lo que uno *sabe* (la moneda es circular, la mesa es rectangular); así el pintor puede observar los verdaderos ángulos subtendidos por las líneas del dibujo (la forma de la moneda es elíptica, la de la mesa es trapezoidal). A veces ayuda el superponer una cuadrícula o un par de hilos cruzados en un visor, para determinar los ángulos de las líneas visibles. Aprender a dibujar consiste en gran medida en aprender a superar los procesos normales de la visión a fin de convertir la experiencia de las cosas vistas en algo *más próximo a mirar una imagen*. Nunca es exactamente como mirar una imagen, pero una vez la cosa vista se ha adulterado lo suficiente en la dirección deseada, es posible, utilizando algunos trucos habituales, «copiar» más o menos lo que uno experimenta sobre el papel.

De forma impresionista, el campo visual nos parece algo uniformemente detallado y enfocado desde el centro hasta las lindes. No obstante, un simple experimento demuestra que no es así. Tome una baraja y coja al azar una carta; manténgala boca abajo para no saber qué carta es. Sitúela entonces en la linde derecha o izquierda de su campo visual y vuélvala, cuidando de mantener la vista al frente (escoja un punto donde fijarla y no deje de mirar en esa dirección). Verá entonces que no es capaz de decir de qué palo es la carta, ni siquiera si es o no es una figura. Fíjese, sin embargo, que es usted capaz de percibir el más pequeño movimiento de la carta. Es usted capaz de percibir el movimiento sin ser capaz de ver la forma o el color de aquello que se mueve. Empiece ahora a mover la carta hacia el centro de su campo visual, siempre con cuidado de mirar hacia el frente: ¿en qué punto puede usted identificar el color de la carta? ¿Cuándo el palo y el número? Observe que usted podrá ver si es una figura antes de poder decir si es caballo, sota o rey. Probablemente, le sorprenderá descubrir lo cerca del centro que puede usted mover la carta antes de poder identificarla por completo.

Por lo general, permanecemos ajenos a esta sorprendente deficiencia de

nuestra visión periférica (toda la visión, con la excepción de un área de dos o tres grados alrededor del centro del campo) a causa del hecho de que nuestros ojos, a diferencia de las cámaras de televisión, no permanecen quietos ante los acontecimientos que el mundo les presenta. Todo lo contrario, no paran de moverse: sin que nosotros nos demos cuenta, llevan a cabo un juego incansable de etiquetado visual de todos aquellos objetos que penetran en nuestro campo de visión y que pueden tener algún interés. Sea a través de un seguimiento continuado, sea por saltos bruscos, nuestros ojos proporcionan a nuestro cerebro información de alta resolución sobre todo lo que ocupa en cualquier momento el centro del área foveal del campo retinal. (La fóvea del ojo tiene una capacidad discriminadora diez veces superior a la de el resto de la retina.)

Nuestra fenomenología visual, los *contenidos* de nuestra experiencia visual, poseen un formato muy diferente de cualquier otro método de representación: ni dibujos, ni películas, ni frases, ni mapas, ni modelos a escala, ni diagramas. Considere lo que está presente en su experiencia cuando usted mira el bullicio de miles de espectadores en un estadio. Los individuos están demasiado lejos para que usted pueda identificarlos, a menos que haya algo lo bastante llamativo como para facilitarle el trabajo (¿el presidente? Sí, claro, seguro que es aquél; es el que está ahí, en medio de todas esas banderolas rojas, azules y blancas). Visualmente, usted puede decir que la multitud está compuesta por seres humanos porque éstos se mueven como se mueve la gente. Hay algo de global en su experiencia visual de la multitud (lo que se ve allí tiene un aire de muchedumbre, de la misma manera que un trozo de árbol visto a través de una ventana tiene aire de olmo o que un suelo puede verse polvoriento); sin embargo, usted no ve solamente una gran mancha de algún modo identificada como «multitud»; usted ve —todo a la vez— miles de detalles particulares: gorritas rojas que se balancean y gafas de sol que brillan, pedacitos de cazadora azul, programas de mano agitados al viento y puños alzados. Si intentáramos pintar una reproducción «impresionista» de su experiencia, esa cacofonía de manchas de colores no capturaría el contenido; usted no tiene la experiencia de una cacofonía de manchas de colores, como tampoco tiene la experiencia de una elipse cuando mira oblicuamente una moneda. Los cuadros —imágenes coloreadas en dos dimensiones— son capaces de reproducir aproximadamente el estímulo que recibe la retina de una escena en tres dimensiones, y crear así una impresión que se asemeja a la impresión visual que se tendría al contemplar directamente la escena, pero el cuadro no es un dibujo *de* la impresión resultante, sino más bien algo que puede provocar o estimular tal impresión.

Es tan imposible pintar una imagen realista de la fenomenología visual, como lo es de la justicia, de una melodía o de la felicidad. Sin embargo, a menudo parece apropiado, incluso irresistible, hablar de las experiencias visuales de uno como si fueran imágenes en la cabeza. Todo ello forma parte del funcionamiento de nuestra fenomenología visual y, por tanto, forma también parte de lo que debemos explicar en los capítulos siguientes.

3. Nuestra experiencia del mundo interior

¿Cuáles son las «materias primas» de nuestras vidas interiores? ¿Y qué hacemos con ellas? Las respuestas no deberían ser demasiado difíciles de encontrar; quizá baste con «mirar y ver» y luego anotar los resultados.

De acuerdo con una tradición todavía tan sólida como la de los empiristas ingleses, Locke, Berkeley y Hume, los sentidos son las puertas por las que se introduce el mobiliario de la mente; una vez dentro, a buen recaudo, este material puede manipularse y combinarse *ad libitum* para crear un mundo interno de objetos imaginados. La manera en que usted se imagina una vaca de color morado volando sería la siguiente: toma usted el color morado, que conoce por haber visto granos de uva, las alas las consigue gracias a haber visto un águila, y se las pega a una vaca, imagen que usted posee por haber visto alguna vez una vaca. Esto no puede ser del todo correcto. Lo que entra en el ojo es una radiación electromagnética, y *no* por ello pasa a ser utilizable como una paleta de colores para pintar vacas imaginarias. Nuestros órganos sensoriales son bombardeados por energía física en las formas más variadas, energía que en el punto de contacto es «traducida» a impulsos nerviosos que viajan hacia el cerebro. Lo que pasa del exterior al interior no es más que información, y, aunque la recepción de información podría *provocar* la creación de alguna entidad fenomenológica (por hablar de la manera más neutral posible), es difícil creer que la información misma —que no es más que una abstracción concretada dentro de un medio físico modulado— pueda ser la entidad fenomenológica en cuestión. Existe todavía alguna buena razón, sin embargo, para estar de acuerdo con los empiristas ingleses en que de *alguna manera* el mundo interior depende de fuentes sensoriales.

La visión es la modalidad sensorial que nosotros, pensadores humanos, casi siempre destacamos como fuente principal de conocimientos perceptivos, aunque luego recurramos al tacto y al oído para confirmar lo que los ojos nos han dicho. Este hábito nuestro de contemplar todo lo que tiene que ver con la mente a través de la metáfora de la visión (hábito en el que hemos caído por lo menos dos veces sólo en esta frase) es una de las principales fuentes de distorsión y confusión, como veremos inmediatamente. La vista domina hasta tal punto nuestra práctica intelectual, que hallamos muchas dificultades en concebir una alternativa. A fin de hacernos comprender, dibujamos diagramas y esquemas para que «se vea lo que pasa», y si queremos «ver si algo es posible», intentamos imaginarlo en «el ojo de nuestra mente». ¿Habría sido capaz una raza de pensadores ciegos que se basara en el oído de comprender con la ayuda de melodías, tintineos y graznidos en la oreja de la mente, lo mismo que nosotros comprendemos gracias a las «imágenes» mentales?

Incluso los ciegos de nacimiento utilizan un vocabulario visual para describir sus propios procesos de pensamiento. No está claro, sin embargo, hasta qué punto esto es el resultado de haberse visto obligados a ceder ante la pre-

sión de un lenguaje que aprende de personas videntes, o de la validez de una metáfora que son capaces de reconocer a pesar de las diferencias en sus propios procesos de pensamiento, o aun debido al hecho de que hacen aproximadamente un mismo uso de la maquinaria visual del cerebro que las personas videntes, todo ello a pesar de carecer de las vías de entrada normales. La respuesta a estas preguntas sin duda contribuiría a aportar nueva luz sobre la naturaleza de la conciencia humana, ya que uno de sus sellos de distinción es precisamente la preponderancia del decorado visual.

Cuando alguien nos explica algo, es frecuente que demostremos haber comprendido con un «ya veo»; y esto no es una mera metáfora muerta. La naturaleza cuasivisual de la *fenomenología* de la comprensión ha sido ignorada casi por completo por los investigadores en ciencia cognitiva, especialmente en inteligencia artificial, que han intentado desarrollar sistemas de comprensión del lenguaje por ordenador. ¿Por qué le han vuelto la espalda a la fenomenología? Quizás en gran medida a causa de su convencimiento de que la fenomenología, por muy real y fascinante que sea, no es funcional; es una rueda que gira sin llegar nunca a engranar en aquella maquinaria que es relevante para la comprensión.

La fenomenología de diferentes oyentes como respuesta a un mismo enunciado puede variar casi *ad infinitum* sin que apenas pueda apreciarse variación en la comprensión o en la captación del contenido. Considérese la variedad de imágenes mentales que podría provocar en dos personas el escuchar el siguiente enunciado:

«Ayer mi tío despidió a su abogado».

Jim podría empezar con un repaso de los malos tragos que pasó *ayer*, todo mezclado con una mirada fugaz a la parte correspondiente a la relación *tío* (hermano del padre o de la madre; marido de la hermana del padre o de la madre) del diagrama de relaciones familiares, y seguido por algunas imágenes de tribunales y de un anciano montando en cólera. Mientras tanto, quizá, Sally habrá pasado por la palabra «ayer» sin recrear imagen alguna y habrá fijado su atención en algún aspecto del rostro de su tío Bill, al tiempo que se imagina una puerta dando un portazo y la silueta apenas «visible» de una mujer bien vestida, marcada con la etiqueta «abogado», saliendo por ella. Independientemente de su imaginería mental, Jim y Sally han comprendido el enunciado igual de bien, como quedaría confirmado por las respuestas y paráfrasis que darían a una batería de preguntas que se les propusiera a continuación. Además, como señalarían los investigadores con mentalidad más teórica, las imágenes *no pueden* ser la clave para la comprensión porque uno no puede hacer el dibujo de un tío, o de un ayer, o de un despedir, o de un abogado. Los tíos, a diferencia de los payasos o los bomberos, no poseen ninguna característica distintiva que pueda ser representada visualmente, y los ayeres no se parecen a nada. Comprender no es, pues, algo que pueda hacerse a través de un proceso de convertirlo todo a la divi-

sa de las imágenes mentales, a menos que los objetos dibujados se identifiquen mediante etiquetas, en cuyo caso la escritura en estas etiquetas sería un mensaje verbal necesario para la comprensión, lo cual nos devolvería al punto de partida.

Mi *oír* lo que usted dice depende de que usted lo diga a una distancia suficiente de mis oídos y mientras estoy despierto, lo cual garantiza que yo lo pueda oír. Mi *comprender* lo que usted dice depende de muchas cosas, pero no parece depender de ningún elemento identificable de la fenomenología interna; ninguna experiencia consciente garantizará que yo le comprenda o le comprenda mal. El hecho de que Sally se imagine a su tío Bill no impedirá de ninguna manera que ella comprenda que es el tío del hablante, y no el de Sally, quien despidió a su abogado; ella *sabe* lo que quería decir el hablante; se ha visto llevada a evocar una imagen de su tío Bill sin que por ello haya riesgo de provocar una confusión, ya que comprender al hablante no depende en modo alguno de su imaginería.[4]

Así pues, la comprensión no puede ser explicada por el mero recurso de citar las entidades fenomenológicas que la acompañan, lo cual no significa que esa fenomenología no esté ahí. Más concretamente, ello no significa que un modelo de comprensión que permanezca en silencio en todo lo que respecta a la fenomenología será capaz de capturar todas nuestras intuiciones sobre la comprensión. No cabe duda de que la principal causa de escepticismo ante la «comprensión por máquina» del lenguaje natural se debe al hecho de que estos sistemas casi nunca incorporan un espacio de trabajo «visual» en el que analizar la información de entrada. Si así lo hicieran, la sensación de que están realmente comprendiendo lo que procesan se vería notablemente potenciada (independientemente de que sea, como insisten algunos, una mera ilusión). Hoy por hoy, si un ordenador profiere «ya veo lo que quieres decir» como respuesta a un información que le hemos introducido, tenemos una fuerte tentación de tachar esa afirmación como un fraude manifiesto.

La tentación es realmente fuerte. Por ejemplo, es difícil imaginar cómo es posible entender ciertos chistes sin la ayuda de las imágenes mentales. Dos amigos están sentados en un bar bebiendo; el uno mira al otro y le dice: «Tío, me parece que te has pasado, ¡la cara se te está poniendo borrosa!». ¿Acaso no ha utilizado usted una imagen o algún rápido bosquejo para imaginarse el error que el hablante acaba de cometer?* Parece que la experiencia nos proporciona un ejemplo de *lo que se siente cuando se entiende algo*: ahí está usted, ante algo confuso o indescifrable o, cuanto menos, desconocido —algo que de un modo u otro provoca un picor epistémico— cuando, de repente, *¡ajá, ya lo tengo!* La comprensión se despierta y la cosa se trans-

4. El desarrollo clásico de esta idea, junto a nuevos argumentos más o menos efectivos, lo hallamos en las *Investigaciones filosóficas* (1953) de Wittgenstein.

* En inglés el error es más evidente, pues el hablante confunde *blushy* (sonrojado, colorado) con *blurry* (borroso, difuminado). [N. del T.]

forma; se convierte en algo útil, accesible, que cae bajo su control. Antes del tiempo *t*, la cosa era incomprensible; después de ese tiempo *t*, ya es comprensible —un claro cambio de estado que a menudo puede ser medido con precisión en el tiempo, aun cuando sea, insistamos en ello, una transición accesible subjetivamente y descubierta por medio de la introspección—. Como veremos, es un error hacer de esto el modelo de todo proceso de comprensión, aunque no deja de ser cierto que cuando el principio de un proceso de comprensión posee algún tipo de fenomenología (cuando somos conscientes de empezar a comprender algo), ésa es la fenomenología que se asocia al proceso en cuestión.

Algo de verdad debe haber en la idea de la imaginería mental, y si las «imágenes en la cabeza» no es la mejor manera de pensar en ello, entonces es preciso hallar una manera mejor de hacerlo. La imágenes mentales afectan a todas las modalidades de percepción sensorial, no sólo a la visión. Imagine usted «Noche de paz», con cuidado de no tararearla mientras lo hace. ¿«Ha escuchado» no obstante la melodía en el oído de su mente sonando a un tono determinado? Si es usted como yo, sí que la habrá escuchado. Yo no entono muy bien, así que no puedo decir «desde dentro» en qué tono me la he imaginado, pero si alguien se pusiera a tocar «Noche de paz» en un piano, sería capaz de decir, sin ninguna duda, «Sí, ésta es la melodía que me he imaginado» o, por el contrario, algo parecido a «No, me la estaba imaginando unos tres tonos por debajo».[5]

No sólo hablamos con nosotros mismos en silencio, sino que a veces también lo hacemos con un «tono de voz» particular. En otras ocasiones es como

5. Un neurocirujano me contó una vez la operación que realizó a un hombre joven con epilepsia. Como es costumbre en estos casos, el paciente estaba despierto, sólo bajo los efectos de una anestesia local, mientras el cirujano exploraba con sumo cuidado su córtex expuesto, asegurándose de que las partes que quizá debían ser extirpadas no eran absolutamente vitales mediante estimulación eléctrica y preguntando al paciente qué sentía. Algunas estimulaciones provocaban destellos visuales o el alzamiento de una mano, otras provocaban como una especie de zumbido, pero hubo un punto que produjo una reacción de placer en el paciente: «Anda, ¡pero si esto es "Outta Get Me", de Guns N' Roses, mi banda favorita de heavy!».

Le pregunté al neurocirujano si le pidió al paciente que cantara o tarareara con la música que oía, pues sería fascinante saber el grado de «alta fidelidad» que alcanzó el recuerdo provocado. ¿Sería exactamente en el mismo tono y *tempo* que en el disco? Este tipo de canciones (a diferencia de «Noche de paz») tienen una versión canónica, así que podríamos haber superpuesto una grabación del paciente cantando a la grabación estándar para comparar los resultados. Desgraciadamente, aunque hubiera habido una grabadora en marcha durante la operación, el cirujano no le pidió al paciente que cantara. «¿Por qué?» le pregunté, y él me contestó: «¡No soporto la música rock!».

Más adelante, durante la misma conversación, el neurocirujano comentó que tenía que volver a operar al mismo paciente, con lo que me apresuré a expresarle mis esperanzas de que en esta ocasión comprobara si podía volver a estimular la música rock y le pidiera al paciente que cantara. «No puedo hacerlo», me contestó el neurocirujano, «porque extirpé aquella parte.» «¿Era parte del foco epiléptico?», le pregunté, y él me contestó, «No, ya te lo he dicho, no soporto la música rock».

La técnica quirúrgica a la que aquí se hace referencia fue desarrollada por Wilder Penfield hace muchos años y se describe muy gráficamente en el libro de Penfield *The Excitable Cortex in Conscious Man* (1958).

si hubiera palabras, pero no palabras oídas, y aún en otras ocasiones, sólo la sombra o el indicio de una palabra está «ahí» para abrigar nuestros pensamientos. En el apogeo de la psicología introspeccionista, se produjeron ardorosos debates sobre si era posible que hubiera pensamientos *totalmente* «sin imágenes». Por el momento, dejaremos abierta esta cuestión; nos limitaremos a señalar que muchos afirman con toda seguridad que los hay, mientras otros afirman con la misma seguridad que no los hay. En el próximo capítulo desarrollaremos un método para tratar con estos conflictos. En cualquier caso, la fenomenología del pensamiento vivo no se limita a *hablar* con uno mismo; podemos hacernos dibujos en los ojos de nuestra mente, conducir un coche con cambio manual, tocar seda, o saborear un imaginario bocadillo de mantequilla de cacahuete.

Tanto si los empiristas ingleses estaban en lo cierto al pensar que estas sensaciones solamente imaginadas (o memorizadas) no eran más que burdas copias de las sensaciones originales que «venían del exterior» como si no lo estaban, no es menos cierto que éstas pueden provocar tanto placer o sufrimiento como las sensaciones «reales». Como todo buen soñador sabe, quizá las fantasías eróticas no sean un sustituto totalmente satisfactorio de la realidad, pero son algo que uno con toda seguridad echaría de menos, si se le impidiera tenerlas. Pues no sólo producen placer; pueden provocar sensaciones reales y algunos otros efectos corporales bien conocidos. Podemos llorar cuando leemos una novela triste, y también puede que haya llorado el novelista en el momento de escribirla.

Todos somos conocedores de los placeres y los dolores de la imaginación, y muchos de nosotros nos consideramos expertos en preparar esos episodios que tanto nos gustan; sin embargo, aún podemos sorprendernos al descubrir cuán poderosa puede volverse esta facultad tras un período serio de aprendizaje. Encuentro particularmente sorprendente, por ejemplo, que cuando se llevan a cabo competiciones de composición musical, los participantes a veces no pesenten grabaciones (o actuaciones en vivo) de sus trabajos; presentan partituras, y los jueces efectúan sus juicios *estéticos* con toda confianza a partir de la lectura de esas partituras y *escuchando la música en sus mentes*. ¿Qué valor tienen las mejores imaginaciones musicales? ¿Puede un músico experimentado, después de leer rápidamente una partitura, saber exactamente *qué* sonido producirán esos oboes y flautas disonantes en relación a la masa sonora de las cuerdas? Anécdotas hay muchas, pero por lo que yo sé éste es un territorio relativamente inexplorado, a la espera de que algún investigador inteligente se interese por él.

Las sensaciones imaginadas (si es que podemos llamar así a estas entidades fenomenológicas) son objetos apropiados para la apreciación y el juicio estéticos, pero, entonces, ¿por qué las sensaciones reales son mucho más importantes? ¿Por qué no nos contentamos con puestas de sol recordadas o con la promesa de unos espaguetis *al pesto*? Gran parte del placer y el dolor que asociamos a acontecimientos de nuestra vida están, después de todo, ligados a la anticipación y al recuerdo. Los simples momentos de sensación

no son más que una pequeña parte de lo que nos interesa. El porqué —y el cómo— del interés que sentimos por las cosas es asunto que trataremos en próximos capítulos, pero el hecho de que las sensaciones imaginadas, anticipadas o recordadas sean tan distintas de las sensaciones *débiles* puede recalcarse recurriendo a otro experimento mental, que nos conduce hasta la puerta de la tercera parte del fenome.

4. Afecto

Ahora cierre los ojos e imagine que alguien le ha dado una patada muy, pero que muy fuerte en la espinilla (unos treinta centímetros por encima del pie) con una bota de puntera metálica. Piense en ese agudísimo dolor con tanto detalle como pueda; imagine que se le saltan las lágrimas, que casi se desmaya, se marea, de tan penetrante y abrumador que es el dolor que usted siente. ¿Se lo ha imaginado bien? ¿Ha sentido el dolor? ¿Podría incluso quejarse porque el seguir fielmente mis instrucciones le ha producido algún dolor? He comprobado que la gente reacciona de maneras muy distintas a este ejercicio, pero nadie hasta ahora me ha dicho que haya sentido ningún dolor. Algunos lo encuentran francamente inquietante, para otros es un ejercicio mental bastante divertido, en cualquier caso nunca tan desagradable como el más leve de los pinchazos en un brazo que usted llamaría dolor.

Suponga ahora que ha soñado la escena de la patada en la espinilla. Un sueño así puede resultar tan chocante que usted puede llegar a despertarse; incluso puede encontrarse agarrándose la espinilla, gimiendo y con lágrimas en los ojos. Pero no habría inflamación ni verdugón ni cardenal, y tan pronto como estuviera usted lo bastante despierto y hubiera recuperado el sentido de la orientación como para fiarse de su propio juicio, diría que ya no había rastro del dolor en la espinilla —si es que lo hubo en algún momento—. ¿Son experiencias los dolores soñados? ¿O son una especie de dolores imaginarios? ¿O algo intermedio? ¿Qué podemos decir de los dolores inducidos por hipnosis?

En cualquier caso, los dolores soñados y los dolores inducidos por hipnosis son estados mentales que nos afectan de algún modo. Compáreselos, sin embargo, con los estados (¿mentales?) provocados durante el sueño cuando uno se mueve, adopta una posición incómoda y, acto seguido, sin despertarse, sin darse cuenta, vuelve a moverse hasta adoptar una posición más cómoda. ¿Son dolores? Si estuviéramos despiertos, los estados causados por la mala postura serían dolores. Hay personas, afortunadamente bastante pocas, que son congénitamente insensibles al dolor. Antes de que empiece a envidiar a estas personas, debería usted saber que como no realizan estos cambios de postura cuando duermen (ni tampoco cuando están despiertos) pronto se convierten en unos lisiados por haber forzado excesivamente las articulaciones a causa de carecer de la señal de alarma apropiada. También

se queman, se cortan y se provocan otras lesiones que van acortando sus infelices vidas por ese mantenimiento inadecuado de sus cuerpos (Cohen y otros, 1955; Kirman y otros, 1968).

No cabe duda de que la posesión de un sistema de fibras nerviosas para el dolor con las regiones asociadas en el cerebro es un adelanto evolutivo, aun si ello significa pagar el precio de tener que soportar *algunas* alarmas que escapan totalmente a nuestro control.[6] Pero, ¿por qué los dolores tienen que *doler* tanto? ¿No sería suficiente con un timbre muy fuerte en el oído de la mente, por ejemplo?

¿Y para qué sirven la ira, el miedo o el odio? (Asumo que la utilidad evolutiva del deseo sexual no necesita defensa.) O, por tomar un ejemplo más complicado, considérese la simpatía. Etimológicamente, la palabra significa *sufrir-con*. Las palabras alemanas correspondientes son *Mitleid* («compasión, piedad», literalmente: con-dolor) y *Mitgefühl* («simpatía», literalmente: con-sentimiento). O piénsese en la llamada *vibración por simpatía*, en la que la cuerda de un instrumento musical empieza a vibrar por la simple presencia de otra cuerda que vibra en su proximidad y que está estrechamente relacionada con ella por compartir la misma frecuencia de resonancia. Suponga que usted presencia una escena en que su hijo sufre algún tipo de humillación o profunda depresión; usted apenas puede soportarlo: olas de profunda emoción le invaden y ahogan sus pensamientos, pierde usted la compostura y siente deseos de luchar, de llorar, de golpear algo. Éste es un caso extremo de simpatía.* ¿Por qué motivo estamos diseñados de modo que tales sensaciones se produzcan en nuestro interior? ¿Y en qué consisten?

En muchos de los capítulos que siguen nos ocuparemos de la significación adaptativa de los distintos estados afectivos (en caso de que la tengan). De momento, sólo quisiera llamar la atención, durante nuestro pequeño paseo, sobre la innegable importancia del afecto para nuestra convicción de

6. La bibliografía sobre la justificación evolutiva del dolor está plagada de argumentos sorprendentemente miopes. Un autor arguye que no puede existir una explicación evolutiva del dolor porque algunos dolores muy agudos, como el producido por los cálculos biliares, hacen sonar una alarma ante la cual nadie podía hacer nada hasta el desarrollo de la medicina moderna. Ningún hombre de las cavernas obtenía beneficio reproductivo alguno por el dolor provocado por unos cálculos biliares, así que el dolor —o cierto tipo de dolores— es un misterio evolutivo. Lo que ignora este autor es el simple hecho de que a fin de tener un sistema de dolor capaz de advertirnos adecuadamente de crisis tan evitables como una garra o un colmillo clavados en el estómago, es bastante razonable suponer que recibiéramos la bonificación —que sólo seríamos capaces de apreciar mucho más tarde— de un sistema que nos avisara de crisis que no fuéramos capaces de resolver. Por el mismo motivo, hay un buen número de estados internos que hoy en día sería de gran utilidad que nos provocaran dolor (el inicio de un cáncer, por ejemplo), y de los que no somos conscientes, quizá porque nuestro pasado evolutivo no incluía ninguna ventaja de supervivencia que fomentara la aparición de las conexiones necesarias (en caso de que emergiera por mutación).

* Como habrá comprobado el lector, en inglés el campo semántico de *sympathy* abarca lo que en castellano, y en alemán, se expresaría con más precisión con palabras como «compasión» (o *Mitleid*). Difícilmente diríamos que la situación descrita en el texto es un caso de «simpatía». [N. del T.]

que la conciencia es importante. Considérese la *diversión*, por ejemplo. Todos los animales *quieren seguir viviendo* —en cualquier caso luchan con todas sus fuerzas por sobrevivir bajo la mayoría de condiciones—, pero sólo algunas especies nos parecen capaces de *disfrutar de la vida* o de *divertirse*. Nos vienen a la mente retozonas nutrias deslizándose por la nieve, cachorros de león jugando, nuestros perros y gatos, pero no arañas o peces. Los caballos, al menos cuando todavía son potros, parecen sentir una cierta alegría de vivir, pero las vacas o las ovejas normalmente parecen estar bastante aburridas o ser, a lo sumo, indiferentes. ¿Acaso no ha pensado usted alguna vez que es una pena que los pájaros vuelen, porque no parecen en general ser capaces de *apreciar* lo maravilloso de su actividad? El de diversión no es un concepto trivial, pero, en mi opinión, no ha recibido hasta ahora la atención que se merece por parte de ningún filósofo. Tengo la certeza de que nunca llegaremos a tener una explicación completa de la conciencia hasta que hayamos dado cuenta del papel que ésta juega en hacer que nosotros (¿sólo nosotros?) nos divirtamos. ¿Qué preguntas es apropiado hacerse? El siguiente ejemplo nos ayudará a ver con qué tipo de dificultades nos enfrentamos.

Existe una especie de primate en América del Sur, más gregario que la mayoría de mamíferos, que muestra una conducta bastante curiosa. Los miembros de esta especie a menudo se reúnen en grupos, grandes o pequeños, y en el curso del parloteo mutuo típico de estas reuniones, bajo una gran variedad de circunstancias, éstos de repente se ven asaltados por unos ataques que se caracterizan por una respiración involuntaria y convulsiva, una suerte de jadeo ruidoso e incontrolado, mutuamente reforzado por los individuos del grupo, que a veces es tan violento que los deja totalmente indefensos. Lejos de ser desagradables, no obstante, estos ataques parecen ser muy del gusto de los individuos de esta especie, que los buscan y en ocasiones muestran una profunda adicción por ellos.

Quizá tengamos la tentación de pensar que si supiéramos lo que estos individuos sienten en su interior, llegaríamos a comprender esta afición suya tan rara. Si pudiésemos verlo «desde su punto de vista», sabríamos para qué sirve. Sin embargo, en este caso podemos estar seguros de que por mucho que lleguemos a saber, la conducta en cuestión seguirá siendo un misterio, porque ya disponemos de la información que buscábamos: la especie es el *Homo sapiens* (que, evidentemente, vive en América del Sur y también en muchos otros sitios), y la conducta es la risa.[7]

Ningún otro animal hace algo así. Un biólogo que se encontrara ante un fenómeno único como éste debería, en primer lugar, preguntarse *para qué* sirve, y, en caso de no dar con ningún análisis plausible en términos de ven-

7. «¿Qué pensaría un visitante del espacio exterior al ver reír a un ser humano? Le debería parecer realmente espantoso: la visión de esos gestos furiosos, esas extremidades agitándose sin control y el tórax afectado de esas horribles convulsiones» Minsky, 1985, pág. 280. [Pág. 290 de la traducción castellana, Galápago, Buenos Aires, 1986; N. del T.]

tajas biológicas directas, se inclinaría por interpretar esta conducta tan rara e improductiva como el precio pagado por el organismo a cambio de alguna otra ventaja. Pero, ¿qué ventaja? ¿Qué cosa hacemos mejor de como la haríamos en caso de no contar con los mecanismos que trae consigo nuestra tendencia —casi adicción— a la risa, y cuyo precio nos merece la pena pagar? ¿Es la risa una manera de «liberar el estrés» que acumulamos en el curso de los complejos procesos cognitivos que jalonan nuestras vidas socialmente avanzadas? Pero, ¿por qué se necesitan cosas *divertidas* para liberar el estrés? ¿Y por qué no las cosas *verdes* o las cosas *planas*? ¿Por qué no gustamos igualmente de permanecer de pie temblando o eructando, o rascándonos la espalda los unos a los otros, o canturreando, o sonándonos la nariz o lamiéndonos fervorosamente las manos?

Nótese que el *punto de vista interno* es bien conocido y no plantea demasiados problemas. Nos reímos porque las cosas son *divertidas* —y la risa es la conducta *apropiada* ante cosas divertidas, de una manera que lamerse las manos, por ejemplo, no lo es—. Es obvio (*demasiado* obvio, de hecho) el porqué nos reímos. Reímos de gozo, de alegría, de felicidad, y porque hay cosas que simplemente causan hilaridad. Si en algún caso se ha podido explicar algo porque posee una *virtus dormitiva* es éste: nos reímos por la hilaridad que causa el estímulo.[8] Es totalmente cierto; no existe ningún otro motivo por el cual nos reímos, cuando nos reímos de verdad. La hilaridad es la causa constitutiva de la verdadera risa. Del mismo modo que el dolor es la causa constitutiva de un comportamiento doliente no fingido. Y ya que es cierto, no tenemos por qué negarlo.

Pero necesitamos una explicación de la risa que vaya más allá de esta verdad tan evidente, igual que las explicaciones tradicionales del dolor y del comportamiento ante el dolor van más allá de lo que es obvio. Somos capaces de ofrecer una explicación biológica coherente de por qué existe el dolor y una conducta de reacción al mismo (de hecho acabamos de esbozarla); lo que queremos es una explicación en los mismos términos del porqué de la existencia de la hilaridad y la risa.

Lo que sabemos de antemano es que si llegamos a tener tal explicación, ésta no satisfará a todo el mundo. Algunos que se llaman a sí mismos *antirreduccionistas* lamentan que la explicación biológica del dolor y de la conducta ante el dolor *se olvide de lo que hay de doloroso* en el dolor, dejando

8. En la última obra de Molière, la clásica comedia *Le malade imaginaire* (1673), Argan, el hipocondríaco del título, al final resuelve todos sus problemas «convirtiéndose» en médico y así puede tratarse a sí mismo. No le hace falta estudiar, basta con chapurrear un poco el latín. En un examen oral burlesco, se le somete a prueba. «¿Por qué», pregunta el examinador, «el opio hace dormir a la gente?» «Porque posee una *virtus dormitiva* (giro latino que significa "propiedad-de-provocar-el-sueño")», responde el candidato. «*Bene, bene, bene, bene respondere*», entona el coro. ¡Muy bien dicho! ¡Qué profundo! ¡Qué preclaro! Y con un espíritu algo más contemporáneo, nos podríamos preguntar qué tiene Michelle Pfeiffer, que sale tan bien en las fotos. Pues, porque es *fotogénica*. ¡He ahí el porqué! (Siempre me lo había preguntado.) En el capítulo 12 consideramos con más detalle la falta de contenido explicativo que tienen aquellas hipótesis que recurren a una *virtus dormitiva*.

de lado ese «horror intrínseco» del dolor que lo convierte en lo que es. Seguramente pondrán las mismas objeciones a cualquier explicación de la risa que demos: dejaríamos de lado su hilaridad intrínseca. Objeciones como éstas a explicaciones del tipo que hemos mencionado son bastante comunes: «Lo que usted ha explicado es la *conducta* y los *mecanismos* que la acompañan, pero le queda explicar *la cosa en sí*, que es el dolor en todo su horror». Todo esto plantea cuestiones bastante complejas, que consideraremos con detalle en el capítulo 12; por el momento señalaremos que cualquier explicación del dolor que *incluyera* el carácter horrible del mismo está condenada a la circularidad, pues permanecería el residuo de una *virtus dormitiva* que no habría sido eliminado. De igual manera, toda explicación de la risa *tiene que* dejar de lado esa presunta hilaridad intrínseca, el gusto, la gracia, porque su presencia no hará más que retardar el intento de hallar una respuesta.

La fenomenología de la risa está herméticamente sellada: *simplemente vemos* directamente, de forma natural, sin inferencias, con una evidencia que va más allá de la «intuición», que la risa es lo que va con la hilaridad, es la reacción «correcta» al humor. Quizá podamos descomponer un poco el fenómeno: la reacción correcta a algo gracioso es la diversión (un estado mental interno); la expresión natural de la diversión (cuando no es preciso ocultarla o suprimirla, como ocurre a veces) es la risa. Ahora parece que tenemos algo que los científicos llamarían una variable intermedia, la diversión, entre el estímulo y la respuesta: y parece además que está constitutivamente ligada a ambos extremos de la cadena. Es decir, la diversión es lo-que-por-definición-provoca-una-risa-sincera, y también es lo-que-por-definición-es-provocado-por-algo-gracioso. Todo esto es evidente. No parece que debamos ir más allá en nuestras explicaciones. Como decía Wittgenstein, las explicaciones tienen que pararse en alguna parte. Pero lo que tenemos aquí no es más que un hecho en bruto —aunque claramente explicable— de la psicología humana. Hemos de ir más allá de la pura fenomenología, si queremos explicar cualquiera de los moradores de nuestro jardín fenomenológico.

Todos estos ejemplos de fenomenología, en su diversidad, parecen tener dos características comunes. Por un lado, nos son muy familiares; nada hay que podamos conocer mejor que las entidades de nuestras fenomenologías personales —eso parece, por lo menos—. Por otro lado, se resisten duramente a un análisis científico materialista; nada puede parecerse menos a un electrón, a una molécula o a una neurona, que *las sensaciones que me produce una puesta de sol en este momento* —eso parece, por lo menos—. Los filósofos siempre se han sentido profundamente impresionados por ambas características, y han desarrollado diversos modos de resaltar lo que para ellos es problemático. Para algunos, el gran enigma reside en esa especial intimidad: ¿Cómo podemos *no equivocarnos nunca* o tener *un acceso privilegiado* o *una aprehensión directa* de estas entidades? ¿Cuál es la diferencia entre las relaciones epistémicas con nuestra fenomenología y nuestras relaciones epistémicas con los objetos del mundo exterior? Para otros, el gran enigma tiene que ver con las «cualidades intrínsecas» poco usuales —o, por

usar la voz latina, los *qualia*— de nuestra fenomenología: ¿Cómo es posible que algo compuesto de partículas materiales *sea* la diversión que estoy sintiendo, o posea la «homogeneidad última» (Sellars, 1963) del cubito de hielo rosa que me estoy imaginando, o pueda llegar a importarme tanto como me importa el dolor que siento?

Hallar una explicación materialista que les haga justicia a todos estos fenómenos no va a ser fácil. Hemos hecho algunos progresos, sin embargo. Nuestro breve inventario ha incluido algunos casos en los que un cierto conocimiento de los mecanismos subyacentes constituye un desafío —quizás incluso una usurpación— a la autoridad que normalmente se confiere a la introspección. Acercándonos un poco más de lo habitual a las jaulas y observando desde diferentes perspectivas a sus moradores, hemos empezado a romper el hechizo, a disipar la «magia» del jardín fenomenológico.

Capítulo 4

UN MÉTODO PARA LA FENOMENOLOGÍA

1. La primera persona del plural

No se puede hacer zoología en serio con sólo pasearse por un zoo, fijándose en esto o en aquello y maravillándose ante las curiosidades con que uno se encuentra. Para hacer zoología en serio se necesita precisión, lo cual depende de disponer de unos métodos de descripción y de análisis compartidos a fin de que los demás miembros de la comunidad de zoólogos puedan comprender lo que uno dice. Para hacer fenomenología en serio, el disponer de un método de descripción claro y neutral es aún más necesario, porque, según parece, no hay dos personas que utilicen las mismas palabras de la misma manera y todo el mundo parece un experto. Es bastante sorprendente el ver con qué frecuencia las discusiones «académicas» en las que se debaten controversias fenomenológicas degeneran en ruidosas cacofonías de puñetazos sobre la mesa, donde nadie escucha a nadie. Y es sorprendente porque, de hecho, según la más firmemente establecida tradición filosófica, *todos estamos de acuerdo* sobre qué es lo que encontramos cuando «miramos dentro» de nuestra propia fenomenología.

Hacer fenomenología siempre se ha presentado como una tarea fiable, consistente en poner en común unas observaciones compartidas por todo el mundo. Cuando Descartes escribió sus *Meditaciones* como un soliloquio en primera persona del singular, claramente esperaba que sus lectores coincidieran con él en cada una de sus observaciones, llevando a cabo los mismos experimentos mentales que él describía y obteniendo los mismos resultados. Asimismo, los empiristas ingleses, Locke, Berkeley y Hume, escribieron con la presunción de que lo que estaban haciendo era, en gran medida, *introspección*, y que sus introspecciones podrían ser fácilmente replicadas por sus lectores. Locke encarnó esta presunción en su *Ensayo sobre el entendimiento humano* (1690) al denominar su método como «método histórico simple». Para Locke éste no debe contener ni deducciones abstrusas ni teorización *a priori*; basta con establecer los hechos observados, recordando a los lectores lo que es manifiesto para todos los que miren allí donde tienen que mirar. De hecho, prácticamente todo autor que ha escrito sobre la conciencia ha caído en lo que podríamos denominar la *presunción de la primera persona del plural*: sean cuales sean los misterios que esconde la conciencia, *nosotros* (usted, mi querido lector, y yo) podemos hablar tran-

quilamente sobre conocidos mutuos, aquellos con los que nos encontramos en nuestras respectivas corrientes de la conciencia. Con la excepción de algunas voces rebeldes, los lectores siempre han sido cómplices de esta conspiración.

Todo esto estaría muy bien si no fuera por el hecho tan embarazoso de que la controversia y las contradicciones complican cualquier afirmación que pueda hacerse bajo estas condiciones de acuerdo mutuo y cordial. Hay algo sobre lo que nos estamos engañando. Quizá nos engañamos al pensar que básicamente somos todos iguales. Quizá cuando las personas tienen conocimiento de las diversas escuelas de pensamiento sobre la fenomenología, se afilian a aquella que les parece mejor, de modo que la descripción que hace cada escuela fenomenológica es esencialmente correcta en lo que respecta al tipo de vida interior de sus miembros. Luego, ingenuamente, se sobregeneraliza, haciendo afirmaciones sobre cómo son los demás que carecen de fundamento.

O quizá nos engañamos al conceder tanto predicamento a la introspección, a nuestra capacidad de observar nuestras propias mentes conscientes. Ya desde Descartes y su *cogito ergo sum*, esta capacidad nuestra ha sido considerada como algo totalmente libre de error; tenemos un acceso privilegiado a nuestros propios pensamientos y sentimientos, un acceso que se nos garantiza que es mejor que el que pueda tener cualquier observador externo. («¡Imagine que alguien intenta convencerle de que usted está equivocado sobre la naturaleza de lo que usted piensa y siente!») Somos «infalibles» —tenemos la seguridad de tener siempre la razón— o como mínimo «incorregibles» —tengamos razón o no, nadie puede corregirnos (Rorty, 1970).

Pero, quizás enmendarnos esta doctrina de la infalibilidad, por muy arraigada que esté, es un error. Quizás, incluso si todos *somos* esencialmente iguales por lo que a nuestra fenomenología respecta, algunos observadores se equivocan cuando intentan describirla, pero al estar tan seguros de tener razón, se muestran relativamente invulnerables a la enmienda. (Son incorregibles en el sentido derogatorio del término.) Sea como sea, se suscitan controversias. Pero existe aún otra posibilidad que, a mi modo de ver, se ajusta más a la verdad: nos estamos engañando al creer que la actividad de la «introspección» *siempre* se reduce a un mero acto de «mirar y ver». Sospecho que cuando afirmamos estar utilizando nuestros poderes de *observación* interna, no estamos haciendo otra cosa que llevar a cabo un acto improvisado de *teorización*; y como teóricos somos bastante simples, precisamente porque hay muy *poco* que «observar» y mucho sobre lo que pontificar sin miedo a caer en contradicciones. Cuando llevamos a cabo un acto de introspección colectivo, nuestra situación se parece mucho a la de los ciegos de la fábula que examinaban las diferentes partes de un elefante. A primera vista puede parecer una idea descabellada, pero veamos hasta dónde nos lleva.

¿Hay algo de lo que encontró en nuestra visita al fenome en el capítulo anterior que le sorprendiera? ¿Acaso le sorprendió, por ejemplo, el no poder identificar el palo del naipe hasta el momento en que éste estaba justo

enfrente de usted? He comprobado que la mayoría de personas se sorprenden, incluso aquellas que saben de las limitaciones de nuestra visión periférica. Si se sorprendió, ello significa que, en caso de que usted hubiera tenido que hablar sobre el tema antes del sorprendente experimento, seguramente habría hecho una descripción incorrecta de los hechos. Con frecuencia, la gente afirma poseer un conocimiento directo del *contenido* de su campo visual periférico del que realmente tienen. ¿Por qué se hacen estas afirmaciones? No es porque las personas hayan podido constatar directa e incorregiblemente que efectivamente poseen contenidos periféricos de esta naturaleza, sino porque eso es lo que *les parece más razonable.* Después de todo, bajo condiciones normales no observamos lagunas en nuestro campo visual, y es seguro que si hubiera un área que claramente no aparece coloreada, percibiríamos la discrepancia; además, miremos donde miremos, todo se nos presenta en color y con todo detalle. Si usted piensa que su campo visual subjetivo es, básicamente, una imagen interna compuesta por formas coloreadas, entonces es razonable que cada porción de la trama sea de *algún* color —¡incluso un lienzo sin pintar es de *algún* color! Pero esto no es más que una conclusión fruto de un modelo bastante cuestionable de su campo visual subjetivo, no es algo que usted pueda observar directamente.

¿Acaso estoy afirmando que no tenemos un acceso privilegiado a nuestra experiencia consciente? No, pero sí estoy diciendo que tendemos a pensar que somos mucho más inmunes al error de lo que realmente somos. Generalmente, cuando se la enfrenta a los defectos del supuesto acceso privilegiado, la gente admite no tener acceso a *las causas y los efectos* de sus experiencias conscientes. Por ejemplo, suelen sorprenderse al constatar que saborean con la nariz y perciben notas muy graves con los pies, pero nunca mantienen poseer una autoridad sobre las causas y los orígenes de sus experiencias. Afirman poseer autoridad, en cambio, sólo sobre las experiencias mismas, aisladas de sus causas y sus efectos. Sin embargo, aunque *digan* poseer autoridad sólo sobre los contenidos aislados de sus experiencias, pero no sobre sus causas y efectos, con frecuencia trascienden las limitaciones que ellos mismos se han impuesto. Por ejemplo, ¿se atrevería a apostar por las proposiciones siguientes? (he inventado como mínimo una):

1. Se puede percibir una mancha que sea roja y verde a la vez —una mancha que es de los dos colores (sin mezclas) a la vez.
2. Si se mira a un círculo amarillo sobre fondo azul (con buena luz), y la luminancia o brillo del amarillo y el azul se ajusta hasta ser igual para ambos colores, la frontera entre el azul y el amarillo desaparece.
3. Existe un sonido, al que a veces se denomina polo auditivo del barbero, cuyo tono parece estar elevándose constantemente, sin llegar a hacerlo verdaderamente.
4. Existe una hierba que, consumida en dosis excesivas, produce el efecto de incapacitar al que la ha tomado para la comprensión oral de frases en su lengua materna. Mientras persisten sus efectos, el oído no

se ve afectado, ni se difuminan los sonidos ni se oyen sonidos adicionales, pero las palabras que se escuchan suenan como si de una lengua extranjera se tratara, aunque el oyente afectado sabe que no es realmente así.

5. Si se cubren los ojos con un antifaz y se aplica un vibrador en un punto del brazo al tiempo que se toca uno la nariz, se tiene la sensación de que ésta crece como la de Pinocho; si se desplaza el vibrador a otro punto, se tendrá la extraña sensación de estar empujando la nariz hacia dentro, como si el dedo índice penetrara hasta el interior del cráneo.

De hecho, he inventado la número 4, pero por lo que sé podría ser perfectamente posible. Después de todo, en una neuropatología muy bien estudiada, la prosopagnosia, la visión del sujeto permanece intacta y éste puede fácilmente identificar objetos con la vista, pero las caras de sus amigos y colegas más cercanos le resultan irreconocibles por completo.[1] Quiero volver a incidir en el hecho de que no es que no poseamos un acceso privilegiado a la naturaleza o al contenido de nuestra experiencia consciente, sino que debemos permanecer atentos a no caer en un muy tentador exceso de confianza en lo que a ello se refiere.

Durante nuestra visita guiada por el fenome, propuse algunos simples experimentos que usted mismo puede llevar a cabo, lo cual ya no se corresponde con el espíritu de la fenomenología «pura». Los fenomenólogos tienden a argumentar que *dado* que no disponemos de autoridad sobre las causas y los efectos fisiológicos de nuestra fenomenología, debemos ignorar tales causas y efectos en un intento de efectuar una descripción pura, neutral y preteórica de lo que nos viene «dado» en el curso de la experiencia diaria. Puede que sí, pero entonces ¡menuda cantidad de curiosos moradores del fenome que nunca llegaríamos a conocer! Es muy probable que a un zoólogo que intentara extrapolar toda una ciencia a partir de la observación de un perro, un gato, un caballo, un petirrojo y un pez de colores se le escapara alguna que otra cosa.

2. La perspectiva de la tercera persona

Puesto que vamos a abandonarnos a la fenomenología *impura*, es preciso que seamos más cuidadosos que nunca con las cuestiones de método. La perspectiva que tradicionalmente han adoptado los fenomenólogos es la *perspectiva de la primera persona* de Descartes, en la cual yo describo en un monólogo (que le dejo escuchar a usted) lo que yo encuentro en mi experiencia

1. Para el fenómeno de la mancha roja y verde, véanse Crane y Piantanida (1983) y Hardin (1988); sobre la desaparición de la frontera entre colores, el llamado efecto Liebmann (1927), véase Spillman y Werner (1990); para el polo auditivo del barbero, véase Shepard (1964); sobre el efecto Pinocho, véase Lackner (1988). Sobre la prosopagnosia, pueden consultarse Damasio, Damasio y Van Hoesen (1982); Tranel y Damasio (1988); y Tranel, Damasio y Damasio (1988).

consciente, contando con que *nosotros* estemos de acuerdo. He intentado demostrar, sin embargo, que esa complicidad tan reconfortante fruto de la *perspectiva de la primera persona del plural* no es más que una traicionera fuente de errores. De hecho, en la historia de la psicología es precisamente el creciente reconocimiento de este problema metodológico lo que marcó la caída del introspeccionismo y el subsiguiente nacimiento del conductismo. Los conductistas tuvieron mucho cuidado en evitar toda especulación sobre lo que pudiera estar ocurriendo en mi mente, tu mente o su mente (de *él, ella* o *ello*). Abogaron por la *perspectiva de la tercera persona*, según la cual únicamente los hechos recogidos «desde el exterior» merecen ser considerados como datos. Podemos grabar en vídeo a unas personas y después medir el índice de error que presentan al efectuar una tarea de movimiento corporal, o los tiempos de reacción al presionar un botón o mover una palanca; podemos medir el pulso, las ondas cerebrales, los movimientos oculares o la respuesta galvánica de la piel (la conductividad eléctrica detectada por los «detectores de mentiras»). Podemos abrir los cráneos de los sujetos (quirúrgicamente o con técnicas de escáner cerebral) para ver lo que ocurre en sus *cerebros*, pero no debemos formular *hipótesis* alguna sobre lo que pudiera estar ocurriendo en sus *mentes*, porque eso es algo sobre lo que no podemos obtener datos mediante los métodos verificables intersubjetivamente de las ciencias físicas.

La idea se resume en pocas palabras: dado que nunca podremos «mirar directamente» dentro de las mentes de las personas y sólo podemos creer en su palabra, todos aquellos hechos que tengan algo que ver con los eventos mentales no forman parte del corpus de datos de la ciencia, ya que éstos nunca podrán ser verificados por métodos objetivos. Este escrúpulo *metodológico* —que, por otra parte, es el principio rector de toda la psicología experimental y la neurociencia actuales (y no solamente de la investigación «conductista»)— con demasiada frecuencia se ha visto elevado al rango de principio *ideológico* bajo una u otra de las formas siguientes:

> Los eventos mentales no existen. (¡Y punto! Es lo que se ha venido a llamar «conductismo desnudo/tosco/grosero».)
>
> Los eventos mentales existen, pero carecen por completo de efectos, de modo que la ciencia no puede estudiarlos (epifenomenismo; véase el capítulo 12, sección 5).
>
> Los eventos mentales existen, y tienen efectos, pero esos efectos no pueden ser estudiados por la ciencia, la cual debe contentarse con desarrollar teorías sobre los efectos y procesos «periféricos» o «inferiores». (Esta visión goza de cierto predicamento entre los investigadores del cerebro, especialmente entre aquellos que recelan de los «teóricos». Es, de hecho, una forma de dualismo; aparentemente, estos investigadores están de acuerdo con Descartes en que la mente *no* es el cerebro, y se sienten satisfechos con disponer únicamente de una teoría del cerebro.)

Todas estas visiones nos llevan a conclusiones injustificadas. Aun en el caso de que los eventos mentales no estuvieran entre los *datos* de la ciencia, elllo no significa que no podamos estudiarlos científicamente. Los agujeros negros y los genes no están entre los datos de la ciencia, pero se han desarrollado respetables teorías científicas sobre los mismos. El desafío reside precisamente en construir una teoría de los eventos mentales, utilizando los datos permitidos por el método científico.

Tal teoría sólo podrá construirse a partir del punto de vista de la tercera persona, porque *toda* ciencia se construye desde esta perspectiva. Algunos le dirán que una teoría de la mente consciente como ésta es imposible. Un caso notable es el del filósofo Thomas Nagel, quien ha sostenido las tesis siguientes:

> Hay cosas del mundo, la vida y nosotros mismos que no pueden ser comprendidas adecuadamente desde un punto de vista eminentemente objetivo, por mucho que éste pueda llevar nuestros conocimientos más allá del punto de donde partimos. Hay demasiadas conexiones con puntos de vista particulares, o con tipos de puntos de vista, por lo que todo intento de desarrollar un análisis completo del mundo en términos objetivos que esté totalmente libre de tales perspectivas conducirá inevitablemente a falsas reducciones o a negar por entero la existencia de ciertos fenómenos cuya realidad es irrefutable. (Nagel, 1986, pág. 7.)

Ya veremos. Es prematuro discutir sobre lo que puede y lo que no puede ser explicado por una teoría hasta que no sepamos exactamente qué es lo que dice esa teoría. Pero si debemos prestar la debida atención a una teoría, a la vista de tanto escepticismo, vamos a necesitar un *método* neutral *para describir los datos*, un método que nos evite prejuzgar el problema. Pudiera parecer que tal método no existe, pero sí que existe un método neutral como el que buscamos. Lo describiré primero, para adoptarlo a continuación.

3. El método de la heterofenomenología[2]

El término tiene algo de amenazador; no sólo fenomenología, sino *hetero*fenomenología. ¿Qué puede ser? De hecho es familiar para todos nosotros, científicos y profanos, pero deberemos introducirlo con una prudencia propia de fanáticos, señalando con exactitud lo que presupone y lo que conlleva, ya que constituye un paso teórico de enormes proporciones. Si evitamos la tentación de tomar atajos, seguiremos el camino *neutral* desde la ciencia física objetiva y su insistencia en el punto de vista de la tercera persona, hasta un método de descripción fenomenológica que (en principio) pueda hacer

2. El contenido de esta sección y la siguiente se basa en mis anteriores análisis de los fundamentos metodológicos de la heterofenomenología: Dennett (1978c, 1982a).

justicia a las más privadas e inefables experiencias subjetivas, sin abandonar los escrúpulos metodológicos de la ciencia.

Queremos una teoría de la conciencia, pero no hay aún acuerdo sobre qué entidades poseen una conciencia. ¿La tienen los humanos recién nacidos? ¿Y las ranas? ¿Y qué decir de las ostras, las hormigas, las plantas, los robots, los zombíes...? De momento, tendremos que permanecer neutrales en lo que a estas criaturas se refiere, pero hay una clase de entidades cuya conciencia nadie pone en duda: nuestros semejantes, los seres humanos adultos.

Sin embargo, algunos de estos seres humanos adultos *podrían* ser zombíes, en el sentido «técnico» de los filósofos. Aparentemente, el término *zombie* tiene sus orígenes en la antigua tradición popular haitiana del vudú y se refiere, en este contexto, a un «muerto viviente», castigado por haber cometido algún crimen y condenado a vagar errante, balbuceando y mirando de hito en hito con sus ojos muertos, obedeciendo ciegamente las órdenes de algún sacerdote o chamán del vudú. Todos hemos visto alguna vez zombíes en las películas de terror y son fácilmente distinguibles de las personas normales. (En pocas palabras, los zombíes haitianos no bailan, ni cuentan chistes, ni mantienen animadas discusiones filosóficas, ni se defienden demasiado bien durante una conversación jocosa, y, además, tienen una pinta espantosa.)[3] Los filósofos, no obstante, utilizan el término *zombie* para referirse a una categoría distinta de ser humano imaginario. De común acuerdo, los filósofos consideran que un zombie es, o sería, un ser humano que aun exhibiendo una conducta perfectamente natural —atento, locuaz, vivo—, no es consciente, sino que es una especie de autómata. El punto fundamental del concepto filosófico de zombie es que uno no puede distinguir a un zombie de una persona normal con sólo examinar su comportamiento externo. Como precisamente es este tipo de conducta lo único a lo que tenemos acceso del comportamiento de nuestros semejantes, *algunos de nuestros mejores amigos podrían ser zombíes*. Ésta es, en cualquier caso, la tradición sobre la cual debo permanecer neutral. Así pues, aunque el método que voy a describir no toma partido sobre la *conciencia real* de los seres

3. Hace algunos años, Wade Davis, un joven antropólogo educado en Harvard, anunció haber desvelado el misterio de los zombíes vudú. En su libro *The Serpent and the Rainbow* (1985) describía la poción neurofarmacológica que preparan los practicantes del vudú y que aparentemente tiene la propiedad de hacer caer a un ser humano en un estado parecido a la muerte; después de haber permanecido enterrados vivos durante algunos días, estas pobres personas a veces son exhumadas para poder suministrarles un alucinógeno que provoca desorientación y amnesia. Sea a causa del alucinógeno, sea a causa de las lesiones cerebrales producidas por la falta de oxígeno durante el período en que permanecieron enterrados, estos individuos realmente se convierten en seres errantes como los zombíes de las películas y, ocasionalmente, pueden acabar esclavizados. A causa del carácter un tanto sensacionalista de las afirmaciones de Davis (así como de la película basada en su libro novelado), sus descubrimientos chocaron en un principio con una sólida barrera de escepticismo en algunos medios, pero esas críticas pronto se vieron acalladas tras la aparición de un segundo libro, mucho más serio en el plano científico, *Passage of Darkness: The Ethnobiology of the Haitian Zombie* (1988). Véanse también Booth (1988) y Davis (1988b).

humanos adultos aparentemente normales, éste se concentra en la clase de seres humanos adultos normales, ya que si existen seres que poseen conciencia, son éstos los principales candidatos para ello. Una vez hayamos sentado las bases para una teoría de la conciencia humana, podremos fijar nuestra atención sobre la conciencia de otras especies (si es que la tienen), incluyendo los chimpancés, los delfines, las plantas, los zombíes, los marcianos, y las tostadoras eléctricas (los filósofos a veces dan rienda suelta a su fantasía con sus experimentos mentales).

Los seres humanos adultos (en adelante, las personas) son el objeto de esudio de muchas ciencias. Sus cuerpos son examinados por biólogos, investigadores médicos, nutricionistas e ingenieros (que se plantean preguntas del tipo: ¿A qué velocidad pueden teclear los dedos de los humanos? ¿Cuál es la resistencia a la tensión del cabello humano?). También son el objeto de estudio de psicólogos e investigadores del cerebro, que ponen a las personas, a las que llaman *sujetos*, ante las más variadas situaciones experimentales. En la mayoría de casos, los sujetos deben ser previamente categorizados y preparados. No sólo es preciso saber cuál es su edad, su sexo, si son zurdos o diestros, su nivel de estudios, etc., sino que hay que explicarles *lo que tienen que hacer*. En este punto radica la principal diferencia entre los sujetos humanos y, por ejemplo, los cultivos de virus de un biólogo, las muestras de materiales raros de un ingeniero, las soluciones de un químico, o los gatos, ratas y palomas del investigador interesado en la psicología animal.

Las personas son los únicos objetos de estudio científico cuya preparación típicamente (pero no siempre) comporta la comunicación verbal. En parte, esto tiene que ver con el problema de la ética científica: las personas no pueden ser utilizadas en experimentos sin su consentimiento expreso, y ese consentimiento no se puede obtener sin interacción verbal previa. Sin embargo, desde nuestro punto de vista, es mucho más importante el hecho de que la comunicación verbal se utilice para preparar y delimitar los experimentos. Se pide a los sujetos que lleven a cabo todo tipo de tareas intelectuales, que resuelvan problemas, que identifiquen objetos dispuestos de una manera determinada, que presionen botones, que emitan juicios, etc. La validez de estos experimentos depende en gran medida de que su preparación se lleve a cabo de forma uniforme y precisa. Si, pongamos por caso, las instrucciones se dan en turco a unos sujetos que sólo hablan inglés, el fracaso del experimento está totalmente garantizado. De hecho, se ha podido comprobar que el más pequeño malentendido al interpretar las instrucciones puede poner en peligro el resultado final del experimento, motivo por el cual es particularmente relevante el validar esta práctica de preparar a los sujetos a través de la comunicación verbal.

¿Qué factores intervienen en esta práctica de hablar a los sujetos? Es un elemento de la experimentación psicológica que no podemos eliminar, pero ¿se presupone la conciencia de los sujetos? ¿Acaso los investigadores no acaban hallándose en la misma situación que los introspeccionistas, viéndose obligados a creer en la palabra del sujeto sin poder verificar qué ha com-

prendido realmente? ¿No corremos el riesgo de ser engañados por zombíes, robots u otro tipo de impostores?

Debemos analizar con detalle las características generales de los experimentos con sujetos humanos. Supongamos que, como ocurre habitualmente, se llevan a cabo múltiples grabaciones de todo el experimento: de vídeo, de sonido, electroencefalográficas, etc. Así, nada que no haya quedado registrado en las grabaciones podrá ser considerado como un dato. Centrémonos en el caso de la grabación de los sonidos —principalmente, sonidos orales— proferidos por los sujetos y los investigadores durante el experimento. Habida cuenta de que los sonidos proferidos por los sujetos se producen con medios físicos, éstos serán, en principio, explicables y predecibles en términos de los mismos principios, leyes y modelos que se utilizan para explicar y predecir el ruido y los estallidos de un motor de coche. En contrapartida, dado que los sonidos tienen también una base fisiológica, podríamos acudir a los principios de la fisiología para explicar los sonidos apelando a los fundamentos de esta ciencia, del mismo modo que explicamos los eructos, los ronquidos, los rugidos del estómago o los crujidos de las articulaciones. Sin embargo, los sonidos que más nos interesan son, evidentemente, los sonidos orales y, en particular, un subconjunto de los mismos (ignorando, pues, los eructos, estornudos y bostezos ocasionales que se pudieran producir) que, *aparentemente*, es susceptible de ser analizado en términos lingüísticos y semánticos. No siempre es obvio saber exactamente qué sonidos deben incluirse en este grupo, pero hay una manera de jugar sobre seguro: entregamos copias de las grabaciones a tres mecanógrafos experimentados para que preparen tres *transcripciones* independientes de los datos.

Este paso tan simple tiene, no obstante, importantes implicaciones; con él, pasamos de un mundo —el mundo de los sonidos físicos— a otro: el mundo de las palabras y los significados, de la sintaxis y la semántica. Este paso comporta una reinterpretación radical de los datos, una abstracción desde sus propiedades acústicas y físicas en general hasta cadenas de palabras (aunque conservando todavía una cronología precisa; véase, por ejemplo, Ericsson y Simon, 1984). ¿Qué factores rigen esta reinterpretación? Si bien es posible que existan ciertas relaciones regulares y detectables entre las propiedades físicas de la onda acústica grabada en la cinta y los *fonemas* que los mecanógrafos oyen y después transcriben en forma de palabras, lo que sabemos sobre la naturaleza de estas relaciones no nos permite describirlas con detalle. (Si no fuera así, el problema de construir una máquina que escribiera al dictado estaría resuelto. Se han realizado avances notables en este terreno, pero permanecen todavía algunas dificultades insoslayables.) En espera de que las investigaciones en acústica y fonología sean más completas, podremos seguir confiando en nuestras transcripciones en tanto que reflejo objetivo de los datos, siempre y cuando tomemos algunas precauciones elementales. En primer lugar, el hecho de que sean mecanógrafos quienes preparen las transcripciones (en vez de que lo haga el propio investigador, por ejemplo) garantiza que no se produzcan, voluntaria o involuntariamen-

te, interpretaciones sesgadas. (Los taquígrafos de los juzgados juegan precisamente ese mismo papel de testimonio neutral.) El disponer de tres transcripciones llevadas a cabo independientemente nos proporciona una medida de la objetividad del proceso. Si las grabaciones son buenas, las transcripciones coincidirán palabra por palabra, con la excepción quizás de una fracción mínima de palabras. Cuando no haya coincidencia, siempre podemos eliminar esos datos particulares o, si así lo preferimos, podemos basarnos en la coincidencia de dos de las transcripciones para fijar el resultado final.

Estrictamente hablando, pues, la transcripción del texto no se nos *proporciona* en tanto que dato, ya que, como hemos visto, ésta se crea a través de un proceso de interpretación. Este proceso de interpretación depende de ciertos supuestos relacionados con la lengua que se utiliza y con ciertas intenciones del hablante. Para percibir este punto con mayor claridad, compárese la tarea de los mecanógrafos con la tarea de transcribir grabaciones del canto de los pájaros o del gruñido de los cerdos. Cuando un hablante humado profiere «l'importa si le do ar botón co'la mano ihquierda», nuestros mecanógrafos estarán todos de acuerdo en transcribir la pregunta «¿le importa si le doy al botón con la mano izquierda?». Sin embargo, esto es así porque conocen el castellano y porque la pregunta tiene sentido en el contexto en que fue proferida. Asimismo, si el sujeto *dijera* «ahora el punto se mueve d'ezquierda a lerecha», permitiríamos que los mecanógrafos normalizaran la transcripción en «ahora el punto se mueve de izquierda a derecha». Con la transcripción de cantos de pájaros o de gruñidos de cerdos, esta depuración no es posible, ni lo será hasta que alguien descubra algún día cuáles son las reglas que rigen la producción de estos sonidos y sea capaz de elaborar y codificar un sistema para describirlos.

Sin ningún esfuerzo —involuntariamente, de hecho— «damos un sentido» a la cadena sonora en el proceso de segmentarla en palabras. (Es mejor que permitamos a los mecanógrafos cambiar «d'ezquierda a lerecha» por «de izquierda a derecha», pues lo más probable es que lo hagan sin darse cuenta.) Que bajo circunstancias normales el proceso sea muy fiable y en gran medida imperceptible, no debería oscurecer el hecho de que es un proceso muy complejo incluso cuando se detiene en el reconocimiento de palabras y no llega a una interpretación completa. Cuando el mecanógrafo transcribe «a mi modo de ver, una triste ceguera embargaba mi presentimiento, una atractiva tendencia de anticipación y de afrenta, un haz de confirmaciones anticipativas que revelaban nuevas superficies detrás de las superficies», puede que no tenga ni la más remota idea de lo que significa, pero lo más probable es que esté totalmente seguro de que éstas son las palabras que pronunció el hablante, sea cual sea su significado.

Es posible que tampoco el hablante sepa cuál es el significado de esas palabras. Después de todo, nuestro sujeto *podría* ser un zombie, o un loro disfrazado de persona, o un ordenador con un buen sistema de síntesis del habla. O, sin necesidad de ir tan lejos, puede que el sujeto se haya confundido, o puede hallarse bajo la influencia de alguna teoría mal comprendida,

o puede estar queriendo jugarle una mala pasada al investigador contándole un montón de cosas sin sentido. De momento, estoy diciendo que el proceso de crear la transcripción de un texto a partir de una grabación es neutral en relación a todas estas extrañas posibilidades, si bien parte del supuesto metodológico de que hay un texto que recuperar. Cuando no existe tal texto, lo mejor es tirar a la papelera los datos sobre el sujeto en cuestión y volver a empezar.

Hasta ahora, el método descrito no es ni muy original ni particularmente controvertido. Hemos llegado a la moderada conclusión de que podemos convertir una grabación magnetofónica en un texto sin transgredir los límites de la ciencia. Nos hemos tomado nuestro tiempo en asegurar el resultado final, porque el próximo paso es el que nos concederá la oportunidad de estudiar empíricamente la conciencia, pero es también el que acarreará los mayores obstáculos y confusiones. Debemos ir más allá del texto; debemos interpretarlo como un registro de *actos de habla*; no ya como una mera proferencia o articulación de palabras, sino como aseveraciones, preguntas, respuestas, promesas, comentarios, demandas de aclaraciones, reflexiones en voz alta o recomendaciones dirigidas a uno mismo.

Esta interpretación nos obliga a adoptar lo que yo he denominado *la actitud intencional* (Dennett, 1971, 1978a, 1987a): debemos tratar a ese emisor de sonidos como un agente, como un agente racional, que posee creencias y deseos y otros estados mentales característicos por manifestar *intencionalidad* o «denotabilidad»,* y cuyas acciones pueden ser explicadas (o predichas) a partir del contenido de estos estados. Así pues, los sonidos proferidos deben ser interpretados como cosas que los sujetos *querían decir*, o *proposiciones* que *querían declarar*, por ejemplo, por *motivos* diversos. De hecho, ya partíamos de estos supuestos cuando en el paso anterior nos permitíamos depurar el texto. (Nuestro razonamiento es el siguiente: ¿por qué motivo *querría alguien decir* «d'ezquierda a lerecha»?)

Cualesquiera que sean los peligros que corramos al adoptar la actitud intencional en relación a estas conductas verbales, constituyen el tributo que debemos pagar si queremos tener acceso a toda una serie de tópicos que explotamos al diseñar experimentos. Existen muchos motivos por los cuales nos apetece decir cosas, y es importante excluir algunas de estas cosas mediante un método experimental. A veces, por ejemplo, las personas sienten el deseo de decir ciertas cosas no porque realmente las crean, sino porque creen que es lo que quieren oír los que las escuchan. Por lo general, es importante adoptar las medidas necesarias para reducir al máximo las posibilidades de que tal deseo esté presente y llegue a hacerse efectivo: ex-

* Traduzco por «denotabilidad« la voz inglesa *aboutness*. Creo que el término se justifica por el hecho de que el concepto de «intencionalidad», tal y como se usa en el texto, hace referencia al carácter representacional de las actitudes proposicionales (creencias y deseos). Los rasgos intencionales o representacionales son rasgos *semánticos*: las creencias son ciertas o falsas; entrañan o implican otras creencias, es decir, su intencionalidad debe entenderse como una capacidad para referir (o denotar). [N. del T.]

plicamos a los sujetos que lo que queremos oír es *aquello que ellos creen*, y ponemos especial cuidado en no hacerles saber qué es lo que esperamos que crean. En resumen, hacemos todo lo que podemos por ponerlos en una situación tal que, dados los deseos que les hemos inculcado (el deseo de cooperar, de recibir una gratificación, de comportarse como sujetos modelo), no tengan otra opción mejor que hacer un esfuerzo por decir efectivamente lo que creen.

Hay otra aplicación de la actitud intencional para con nuestros sujetos que es necesaria si queremos valernos de acontecimientos del tipo de presionar un botón. Generalmente, presionar un botón es una manera de llevar a cabo un acto de habla específico como, por ejemplo, *aseverar* que en *este momento*, para mí, ambas figuras aparecen superpuestas, o responder que *sí*, que mi juicio espontáneo, precipitado (porque me han dicho que la rapidez es esencial), es que la palabra que acabo de oír estaba en la lista que escuché hace unos instantes. Por motivos experimentales, por tanto, querremos hacer explícito el significado de actos tales como presionar un botón, a fin de incorporarlo como un elemento más del texto. El tipo de acto de habla a que corresponde cada acción particular de presionar un botón depende de la interpretación intencional que se dé a las interacciones entre sujeto e investigador que se produjeron durante la preparación del sujeto para el experimento. (No toda acción de presionar un botón corresponde necesariamente a un acto de hablar; algunos de estos actos pueden ser, por ejemplo, actos fingidos de disparar o de conducir un cohete.)

Cuando surgen dudas sobre si el sujeto ha dicho lo que quería decir, o ha comprendido el problema, o conoce el significado de las palabras que se han utilizado, siempre se pueden pedir aclaraciones. Normalmente, las dudas llegan a resolverse. En teoría, el efecto que tienen estas medidas es suprimir toda fuente de ambigüedad e incertidumbre de la situación experimental, de modo que *una* de las posibles interpretaciones intencionales del texto (incluidos los actos de presionar un botón) no tenga competidores creíbles. Se convertirá así en *la* expresión sincera y fiable, por parte de un *único sujeto unificado*, de las creencias y opiniones de ese mismo sujeto.[4] Como veremos, sin embargo, hay ocasiones en que tal supuesto es problemático, especialmente cuando nuestros sujetos padecen algún tipo de patología. ¿Qué actitud debemos adoptar ante las personas que, con aparente sinceridad, afirman no ver nada en casos de lo que se ha venido a denominar ceguera histérica, y ante la negación de su ceguera, también aparentemente

4. En «How to change your mind», en *Brainstorms* (1978a), adopto un uso convencional del término «opinión» que me permite establecer una distinción entre las creencias propiamente dichas y otros estados más influidos por el lenguaje, que llamo opiniones. Los animales sin lenguaje pueden tener creencias, pero no opiniones. Las personas pueden tener ambas cosas, pero si usted cree que mañana es viernes, ello deberá ser considerado, de acuerdo con mi definición, su opinión de que mañana es viernes. No es el tipo de estado cognitivo que uno pueda tener sin lenguaje. Aunque no presupongo familiaridad con esta distinción, mi intención es que mis afirmaciones en el texto sean aplicables a ambas categorías.

sincera, que muestran aquellos individuos ciegos que padecen anosognosia (negación de la ceguera o síndrome de Anton)? Nos ocuparemos de estos fenómenos en capítulos posteriores. Si lo que queremos es comprender lo que estas personas experimentan, es evidente que no lo conseguiremos con una simple entrevista.

4. Mundos ficticios y mundos heterofenomenológicos

Además de los problemas particulares que plantean los casos raros, parece existir un problema general. ¿Acaso la práctica de interpretar la conducta verbal no equivale a suponer la conciencia de los sujetos y, por lo tanto, constituye una petición de principio en relación al problema de los zombíes? Suponga que usted se enfrenta a un ordenador «parlante», y suponga que consigue interpretar sus proferencias como actos de habla que expresan sus creencias y opiniones, presumiblemente «acerca de» sus estados conscientes. El hecho de que *haya* una única interpretación coherente de una secuencia de conductas no significa que esa interpretación sea *verdadera*; podría ser que el «sujeto» se estuviera comportando *como si* fuera consciente; nos arriesgamos a vernos engañados por un zombie que carezca por completo de vida interior. Con este método de interpretación, usted no podría *confirmar* si el ordenador era consciente o no. Está claro. No podemos estar seguros de que los actos de habla que observamos expresen verdaderas creencias sobre experiencias reales; quizá solamente expresen creencias *aparentes* sobre experiencias *inexistentes*. Sin embargo, el que hayamos dado con una interpretación estable de la conducta de una entidad determinada en términos de actos de habla siempre será un hecho merecedor de nuestra atención. Cualquiera que descubra una manera intersubjetivamente uniforme de interpretar el movimiento de las ramas de un árbol provocado por la brisa como «comentarios» del «tiempo» sobre la situación política actual, habrá encontrado algo maravilloso que merece ser explicado, aun en el caso de que resultaran ser los efectos de algún artilugio creado por un ingeniero guasón.

Afortunadamente, existe una analogía que nos ayudará a *describir* estos hechos sin que tengamos que *explicarlos* al mismo tiempo: podemos comparar la tarea del heterofenomenólogo de interpretar la conducta de los sujetos en el acto de interpretar una obra de ficción llevado a cabo por un lector. Se sabe (o se asume) que algunos textos, como las novelas y los cuentos, son ficción, lo cual no constituye un obstáculo para su interpretación. De hecho, al eliminar o posponer ciertas cuestiones complejas relacionadas con la sinceridad, la veracidad y la referencia, ello facilita, en determinados aspectos, la tarea de interpretar.

Considérense algunos hechos incontestables sobre la semántica de la ficción (Walton, 1973, 1978; Lewis, 1978; Howell, 1979). Una novela cuenta una historia, pero no una historia verdadera, salvo por accidente. A pesar de que

sepamos o asumamos que la historia que se cuenta no es verdadera, podemos hablar, y de hecho lo hacemos, de lo que es *verdad en la historia*. «Sin faltar a la verdad, podemos decir que Sherlock Holmes vivía en Baker Street y que le gustaba hacer alarde de sus capacidades mentales; en los mismos términos, no podemos decir, sin embargo, que fuera un devoto padre de familia o que trabajara en estrecha colaboración con la policía» (Lewis, 1978, pág. 37). En una historia hay muchas más cosas verdaderas de las que explícitamente se exponen en el texto. Por ejemplo, es cierto que no había aviones a reacción en el Londres de Holmes (aunque ello no se afirme de forma explícita ni sea una consecuencia lógica de lo que se afirma en el texto), pero también es cierto que había afinadores de pianos (aunque —por lo que recuerdo— no se mencione a ninguno, ni tampoco su existencia se deduzca lógicamente de lo dicho). Además de lo que es verdadero y falso en una historia, existe una vasta zona indeterminada: mientras que es verdad que Holmes y Watson tomaron el tren de las 11.10 de la estación de Waterloo a Aldershot en un día de verano, no es ni verdad ni mentira que ese día fuera miércoles (en «The Crooked Man»).

Existen problemas filosóficos deliciosos sobre cómo decir (estrictamente) todo aquello que queremos expresar sin confusiones cuando hablamos de ficción, pero no son éstos los que nos interesan aquí. Puede que haya personas que se sientan intrigadas por la condición metafísica de los objetos y los personajes ficticios, pero yo no soy una de ellas. Desde mi jovial optimismo, no puedo más que suponer que las respuestas que podemos ofrecer, en el plano ontológico, a la cuestión de los resultados de la ficción no plantean ningún problema filosófico serio; la ficción es *ficción*; Sherlock Holmes *no existe*. Dejando de lado, pues, estos detalles y las ingeniosas propuestas técnicas diseñadas para tratarlos, quisiera llamar la atención sobre un hecho muy simple: es innegable que la interpretación de la ficción es factible y que ello puede producir algunos resultados incontestables. En primer lugar, el desarrollo de la historia, la exploración de «el mundo de Sherlock Holmes», por ejemplo no es ni absurdo ni ocioso; se pueden aprender muchas cosas sobre una novela, sobre su sentido, sobre el autor, incluso sobre el mundo real al aprender sobre el *mundo descrito* en la novela. En segundo lugar, si ponemos cierto cuidado en identificar y excluir juicios de valor o preferencias (por ejemplo, «Watson es un pelmazo presumido»), podemos acumular un volumen considerable de hechos incontestablemente objetivos sobre el mundo que se describe. Todos los intérpretes coinciden en afirmar que Holmes era más listo que Watson; la objetividad reside en la obviedad aplastante.

En tercer lugar —y esto es un motivo de alivio para los estudiantes—, el conocimiento del mundo descrito en la novela puede ser independiente del conocimiento que se tenga del texto de la novela. Yo podría llegar a hacer un trabajo aceptable sobre *Madame Bovary*, aunque no he leído la novela (ni siquiera en su traducción inglesa). He visto la serie de televisión realizada por la BBC, así que conozco la historia. Sé lo que ocurre en ese mundo.

Con ello pretendo demostrar, de manera general, que los hechos sobre el mundo de una ficción son hechos a un *nivel* puramente *semántico* de esa ficción; son independientes de los hechos sintácticos sobre el texto (si la ficción es un texto). Podemos comparar las versiones para teatro y cine de *West Side Story* con la obra de Shakespeare *Romeo y Julieta*; al describir las similitudes y las diferencias de lo que ocurre en esos mundos, percibimos similitudes en las obras de arte que no pueden ser descritas en los mismos términos que son apropiados para una descripción sintáctica o textual (y mucho menos física) de reproducciones concretas de esas ficciones. El hecho de que en cada mundo haya una pareja de amantes que pertenecen a ficciones distintas no es un hecho sobre el vocabulario, la estructura oracional, la longitud (en palabras o fotogramas), el tamaño, la forma y el peso de cada una de las reproducciones físicas particulares de esos trabajos.

En general, es posible describir lo que se retrata en una obra de arte (por ejemplo, *Madame Bovary*) independientemente de la descripción que se haga de *cómo* se lleva a cabo ese retrato. (Es evidente que por lo general no se intenta hacer esa separación, sino que se mezclan observaciones sobre el mundo descrito con observaciones sobre los medios que pone en práctica el autor para llevar a cabo la descripción; tal separación no deja de ser posible, sin embargo.) Es incluso posible imaginar una situación en la que se posean los conocimientos suficientes sobre el mundo descrito como para ser capaz de identificar al autor de la ficción, sin ni siquiera conocer el texto o cualquier traducción lo bastante fiel del mismo. Al tener un conocimiento indirecto de lo que ocurre en una ficción, se puede llegar a hacer aseveraciones como: sólo Wodehouse podría haber imaginado tales desventuras. Pensamos que podemos calificar tipos de acontecimientos y circunstancias (y no sólo tipos de *descripciones* de acontecimientos y circunstancias) como kafkianos, y que podemos identificar personajes como típicamente shakespearianos. No cabe duda de que muchas de estas convicciones son erróneas (como es posible demostrar con ciertos experimentos bastante ingeniosos), pero no todas lo son. Las traigo a colación precisamente para ilustrar que podemos conseguir muchas cosas sólo a partir de *lo que se representa*, aun cuando poseamos un conocimiento muy pobre de *cómo se lleva a cabo tal representación*.

Apliquemos ahora la analogía al problema con que se enfrenta el investigador que quiere interpretar los textos producidos por sujetos, sin prejuzgar si los sujetos son zombíes, ordenadores, si están mintiendo o confundidos. Consideremos las ventajas de adoptar la táctica de interpretar estos textos como una especie de ficciones, evidentemente no como literatura, pero sí como generadores de *la ficción de un teórico* (la cual podría resultar ser cierta finalmente). El lector de una novela permite al texto *erigir* un mundo (ficticio), un mundo que el texto estipula por decreto, exhaustivamente extrapolado hasta donde puede llegar la extrapolación, e indeterminado más allá de este punto; nuestro investigador, el heterofenomenólogo, permite al texto del sujeto *constituir* el *mundo heterofenomenológico* del sujeto, un mun-

do que el texto estipula por decreto (en tanto que texto interpretado) y que queda indeterminado más allá. Todo ello hace que el heterofenomenólogo pueda posponer la solución de aquellos problemas espinosos sobre la relación que pudiera existir entre este mundo (ficticio) y el mundo real. Esto hace que los teóricos puedan estar de acuerdo sobre qué *es* el mundo heterofenomenológico de un sujeto, aun cuando su visión sobre cómo se corresponden esos mundos heterofenomenológicos con eventos en el cerebro (o, si se quiere, en el alma) sea totalmente distinta. El mundo heterofenomenológico del sujeto será un postulado teórico estable e intersubjetivamente confirmable, con el mismo estatuto metafísico que, pongamos por caso, el Londres de Sherlock Holmes o el mundo según Garp.

En tanto que ficción, lo que el autor (el autor aparente) dice, vale. Más concretamente, lo que el autor aparente dice proporciona un texto que, cuando es interpretado de acuerdo con las reglas que acabamos de exponer, vale para estipular cómo *es* un «mundo» determinado. No nos preguntamos cómo llegó Conan Doyle a saber de qué color era la butaca de Holmes, ni tampoco se nos ocurre plantear la posibilidad de que se equivocara; es cierto que corregimos errores tipográficos y, por lo demás, hacemos la mejor y más coherente lectura del texto de que somos capaces. De modo similar, no nos preguntamos cómo saben los sujetos (los sujetos aparentes) lo que afirman, y (en este punto) ni siquiera *consideramos* la posibilidad de que estén equivocados; aceptamos su palabra (tal y como nosotros la interpretamos). Nótese, además, que, aunque a veces las novelas incluyan la aclaración de que las descripciones que en ellas se hacen no se refieren a personajes reales, vivos o muertos, la táctica de permitir que el texto erija un mundo no tiene por qué restringirse a obras literarias *concebidas* como ficciones por sus autores; podemos describir a la Reina Victoria de un biógrafo determinado o el mundo de Henry Kissinger, sin atender a las presuntas intenciones del autor de contar la verdad y de referirse, no como un acto fortuito, a personas reales.

5. El discreto encanto de la antropología

Esta manera de tratar a las personas, como si fueran generadores de ficciones (teóricas), no es la manera normal de tratar a la gente. El *conceder* simplemente autoridad constitutiva a sus declaraciones puede resultar un tanto paternalista, como si les mostráramos un falso respeto y no un respeto genuino. Ello es particularmente evidente en una aplicación ligeramente distinta de la táctica heterofenomenológica que llevan a cabo los antropólogos. Un ejemplo nos será útil para clarificar este punto. Supongamos que unos antropólogos descubren una tribu que cree en un dios de los bosques, hasta entonces desconocido, llamado Fenhomo. Después de tener noticia de Fenhomo, los antropólogos se enfrentan a un dilema fundamental: pueden convertirse a la religión nativa y creer fervientemente en la existencia y las buenas obras de Fenhomo, o pueden *estudiar el culto* con una actitud agnós-

tica. Consideremos esta última. A pesar de no creer en Fenhomo, los antropólogos deciden estudiar y sistematizar tan bien como puedan la religión de estas gentes. A partir de lo que les cuentan informantes nativos, establecen descripciones de Fenhomo. Intentan hallar un cierto consenso, pero no siempre lo encuentran (algunos dicen que Fenhomo tiene los ojos azules, otros dicen que los tiene castaños; ni siquiera está claro cuál es su sexo). Intentan explicar y eliminar estas discrepancias, identificando e ignorando la palabra de los sabelotodos, explorando reformulaciones con la ayuda de sus informantes y, quizás, incluso mediando en alguna disputa. Poco a poco va surgiendo un constructo lógico: Fenhomo el dios (o diosa) de los bosques, con una lista exhaustiva de rasgos y hábitos y una biografía. Estos científicos agnósticos (que se hacen llamar a sí mismos fenhominólogos) han descrito, ordenado y catalogado parte del mundo constituido por las creencias de los nativos, y (si han hecho bien su trabajo de interpretación) han compilado una descripción *definitiva* de Fenhomo. Las creencias de los nativos creyentes (que llamaremos fenhoministas) son dogmáticas (se trata de *su* dios, después de todo), pero sólo porque Fenhomo está siendo tratado como un *mero* «objeto intencional», una mera ficción a ojos de los infieles, y por tanto como una *criatura* de las creencias (verdaderas o falsas) de los fenhoministas. Dado que estas creencias pueden contradecirse las unas a las otras, a Fenhomo, en tanto que constructo lógico, se le pueden atribuir propiedades contradictorias; esto es perfectamente posible a ojos de los fenhominólogos, ya que para ellos es *solamente* un constructo. Los fenhominólogos intentan presentar el mejor constructo lógico posible, pero no tienen ninguna obligación de resolver todas las contradicciones que se presenten. Están preparados para detectar desacuerdos entre los devotos que sean imposibles de resolver o de eliminar.

Evidentemente, los fenhoministas no lo ven así, por definición, ya que ellos son los creyentes para los cuales Fenhomo no es un *mero* objeto intencional, sino alguien tan real como usted o como yo. Su actitud para con su propia autoridad sobre la naturaleza de Fenhomo es —o debería ser— un poco más complicada. Por un lado, creen que lo *saben* todo sobre Fenhomo; ellos son, después de todo, fenhoministas, y ¿quién va saber más que ellos? Sin embargo, a menos que se consideren en posesión de algo así como una infalibilidad papal, aceptan la posibilidad de estar equivocados con respecto a ciertos detalles. Es posible, pues, que puedan recibir enseñanzas sobre la verdadera naturaleza de Fenhomo. Por ejemplo, el propio Fenhomo podría ayudarles a corregir algunos detalles. Así pues, deberían sentirse un tanto molestos ante esa credulidad contenida (desde su punto de vista) de los investigadores fenhominólogos, que casi siempre se los toman escrupulosamente al pie de la letra, sin polemizar, ni dudar, siempre planteando respetuosas preguntas para resolver ambigüedades y aparentes conflictos. Un fenhominista nativo que acabara por aceptar la actitud de los antropólogos debería adoptar una postura distante o neutral hacia sus propias convicciones (¿o deberíamos decir sus propias convicciones *anteriores*?), y, durante el proceso, iría abandonando las filas de los verdaderamente devotos.

El método heterofenomenológico ni cuestiona ni acepta como totalmente verdaderos los testimonios de los sujetos, sino que mantiene una neutralidad constructiva y comprensiva, con la esperanza de llegar a compilar una descripción *definitiva* de cómo es el mundo según los sujetos. Todo sujeto que se sintiera incómodo al atribuírsele esa autoridad constructiva podría protestar: «¡No, *de verdad!* ¡Lo que estoy describiendo es *perfectamente real*, y tiene exactamente las propiedades que estoy diciendo que tiene!». La respuesta del heterofenomenólogo honesto sería asentir y asegurar al sujeto que su sinceridad no está siendo puesta en duda. Sin embargo, como los creyentes por lo general quieren más —quieren que se crea lo que dicen o, en su defecto, quieren saber cuándo su audiencia no les cree— suele ser una actitud más diplomática por parte de los heterofenomenólogos, sean antropólogos o investigadores que estudian la conciencia en un laboratorio, el evitar llamar la atención sobre su neutralidad oficial.

Esta desviación de las relaciones interpersonales normales es el precio que se debe pagar en aras de la neutralidad que requiere una ciencia de la conciencia. Oficialmente, debemos mantener una mente abierta ante la posibilidad de que nuestros sujetos aparentes estén mintiendo, sean zombíes o loros disfrazados de personas, pero nunca debemos correr el riesgo de que éstos se molesten haciendo público nuestro descubrimiento. Además, la táctica de la neutralidad no es más que una etapa en el camino de elaborar y confirmar una teoría empírica capaz, en principio, de justificar a los sujetos.

6. A la descubierta de lo que uno está diciendo realmente

¿Qué significaría confirmar las creencias de los sujetos en términos de su propia fenomenología? Con la ayuda de una de nuestras analogías podremos ver mejor las diferentes posibilidades. Considérese cómo podríamos confirmar si una «novela» es una biografía auténtica (o parcialmente auténtica). Podríamos empezar haciéndonos toda una serie de preguntas como ¿en qué persona real dentro del círculo de conocidos del autor se inspira el personaje? ¿Es ese personaje la madre del autor disfrazada? ¿Qué acontecimientos reales de la infancia del autor se han transformado como por encanto en este episodio ficticio? ¿Qué esta intentando decir el autor *realmente*? Preguntar directamente al autor puede no ser la mejor manera de hallar una respuesta para estas preguntas, ya que es posible que ni él las conozca. A veces es razonable suponer que el autor se ha visto obligado, inconscientemente, a expresarse de forma alegórica o metafórica. Los únicos recursos expresivos de que disponía —por el motivo que fuera— no le permitían llevar a cabo una relación directa, objetiva y libre de metáforas de los acontecimientos que quería contar; la historia que ha compuesto es o un compromiso o una cortina de humo. Y como tal puede ser drásticamente reinterpretada (si es necesario, desoyendo las angustiadas protestas del autor), a fin de revelar una historia real sobre personas y acontecimientos reales. Se podría

argumentar que, habida cuenta de que seguramente no es una coincidencia que tal o cual personaje de ficción posea tales o cuales rasgos, podremos entonces reinterpretar el texto que describe este personaje de manera que sus términos denoten —de manera real y no de manera ficticia— los rasgos y los actos de una persona real. Un retrato que nos presente a la Molly ficticia como una prostituta puede sin darse cuenta ser legítimamente considerado como una calumnia por parte de la Polly real, ya que todo lo que hace referencia a Molly se refiere *en realidad* a Polly. Las protestas del autor podrán quizás inducirnos a pensar, correcta o incorrectamente, que la calumnia no era, en ningún caso, un acto consciente o deliberado, pero hace ya tiempo que Freud y otros autores nos convencieron de que los autores, igual que nosotros, a menudo ignoran cuál es la fuente de sus intenciones. Si se puede calumniar sin darse cuenta, necesariamente tales actos deben poder ir acompañados de referencias inconscientes.

O, por utilizar otra de nuestras analogías, considérese qué ocurriría si un antropólogo llegara a confirmar la existencia de un individuo que se llama Fenhomo, que cura a los enfermos y que viaja por la selva de rama en rama como Tarzán. No es un dios, no sabe volar ni tiene el don de la ubicuidad, pero es la incuestionable fuente de todas las visiones, leyendas y creencias de los fenhoministas. Naturalmente, ello causaría una amarga desilusión entre los fieles: algunos abogarían por una revisión atenuada de su credo, mientras que otros se aferrarían a la visión ortodoxa, a pesar de que ello significara reunir al Fenhomo «real» (con sus poderes sobrenaturales intactos) y a su representante, vivo y de carne y hueso, en el mundo. La resistencia que oponen los ortodoxos a la idea de que estaban en un error con respecto a Fenhomo es comprensible. A menos que el candidato de los antropólogos a referente real de la doctrina de los fenhoministas posea un parecido sorprendente, en propiedades y dones, con el Fenhomo de las leyendas, no existirá ningún motivo para anunciar tal descubrimiento. (Compárese con la situación siguiente: «He descubierto que Papá Noel existe. Es un violinista alto y delgado que vive en Miami y que se llama Fred Dudley; odia a los niños y nunca compra regalos».)

Mi sugerencia es, pues, que si llegáramos a encontrar acontecimientos reales en los cerebros de la personas, que poseyeran un número *suficiente* de las propiedades «definitorias» de los elementos que pueblan los mundos heterofenomenológicos, sería razonable pensar que habíamos descubierto lo que querían decir *realmente*, aun cuando en un principio opusieran cierta resistencia a aceptar tales identificaciones. Y si descubriéramos que los acontecimientos reales tienen sólo un pequeño parecido con los elementos heterofenomenológicos, sería razonable declarar que, a pesar de su sinceridad, las personas se equivocan en cuanto a la expresión de sus creencias. Siempre quedará la posibilidad de que alguien —un fenhominista empedernido— insista en afirmar que los elementos fenomenológicos reales *acompañan* a los acontecimientos sin ser idénticos a ellos, lo cual nada tiene que ver con el hecho de que esta afirmación sea convincente o no.

Como los antropólogos, podemos permanecer neutrales al explorar este dominio. Tal neutralidad puede parecer carente de sentido: ¿acaso no es simplemente inimaginable que los científicos lleguen a descubrir fenómenos neurofisiológicos que *sean* los tan celebrados elementos a que hacen referencia los sujetos en sus heterofenomenologías? Los eventos cerebrales parecen diferir demasiado de los elementos fenomenológicos como para ser los referentes reales de las creencias que expresamos en nuestros testimonios introspectivos. (Como vimos en el capítulo 1, parece que necesitamos que la sustancia mental sea la sustancia de que se componen las vacas de color violeta y otras cosas parecidas.) Sospecho que la mayoría de las personas siguen encontrando inimaginable la posibilidad de realizar tal identificación; sin embargo, en vez de aceptar que ello sea realmente imposible, prefiero que sigamos ensanchando nuestra imaginación, con la ayuda de otra fábula. Una fábula que intenta poner el cerco a un elemento fenomenológico particularmente enigmático, la *imagen mental*, y que tiene la virtud de ser en gran parte una historia real, que me he permitido simplificar y adornar un poco.

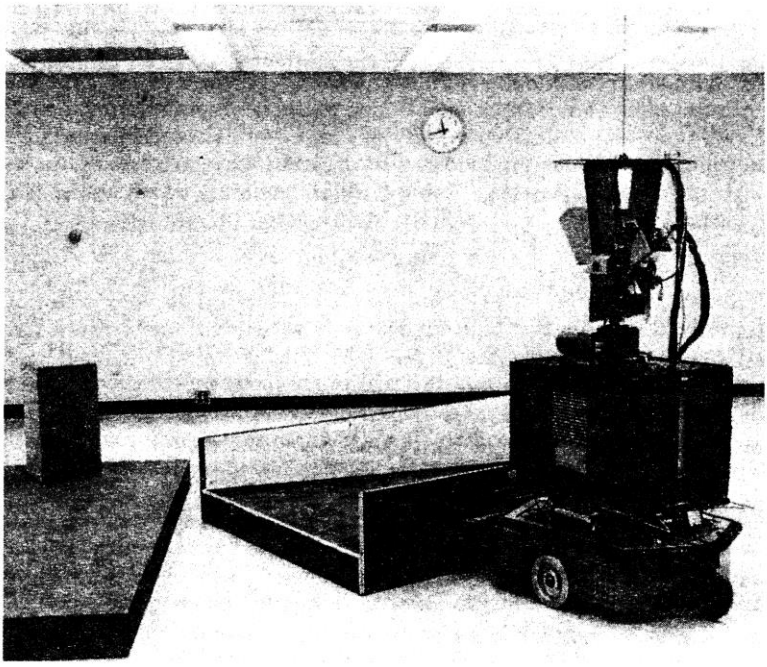


Figura 4.1

7. Las imágenes mentales de Shakey

En la breve historia de los robots, Shakey, construido en el Stanford Research Institute de Menlo Park, California, a finales de los años sesenta por Nils Nilsson, Bertram Raphael y sus colegas, merece ser calificado de leyenda; y no porque desarrollara alguna especial habilidad o porque fuera una simulación particularmente realista de algún rasgo de la psicología humana, sino porque con sus maneras de alienígena abrió ciertas posibilidades intelectuales y cerró otras (Raphael, 1976; Nilsson, 1984). Era el tipo de robot que un filósofo podría admirar, una especie de argumento con ruedas.

Shakey era una caja con ruedas y con una cámara de televisión que funcionaba a modo de ojo, y, en vez de llevar su cerebro consigo, estaba conectado a él (en aquellos tiempos, uno de esos ordenadores estacionarios enormes) por radio. Shakey vivía en unas habitaciones donde no había más que unas cuantas cajas, pirámides, rampas y plataformas, cuidadosamente coloreadas e iluminadas para facilitar la «visión» de Shakey. Para comunicarse con él, había que teclear un mensaje desde un terminal conectado a su cerebro-ordenador, en un pseudoinglés de vocabulario muy restringido. La orden «Push the box off the platform» (literalmente, «empuja la caja fuera de la plataforma») enviaba a Shakey a buscar la caja, localizar una rampa, empujar la rampa hasta ponerla en la posición adecuada, subir por la rampa hasta la plataforma y, finalmente, empujar la caja fuera de la plataforma.

¿Cómo hacía Shakey todo esto? ¿Quizá había un pequeño gnomo en el interior de Shakey, mirando una pantalla de TV y manejando los controles? El recurso a un único e inteligente homúnculo como éste sería una de las maneras de hacerlo —aunque también sería hacer trampa—. Otra manera de conseguirlo consistiría en situar a un controlador humano fuera del cuerpo de Shakey, con un sistema de control remoto por radio. Ésta sería la solución cartesiana, con el transmisor/receptor de Shakey jugando el papel de la glándula pineal, y las señales de radio como dobles —nada milagrosos en este caso— de los mensajes no físicos del alma de que hablaba Descartes

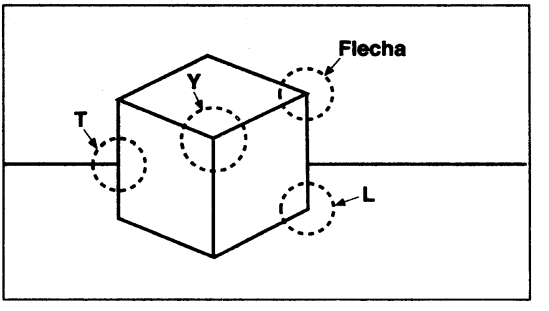


Figura 4.2

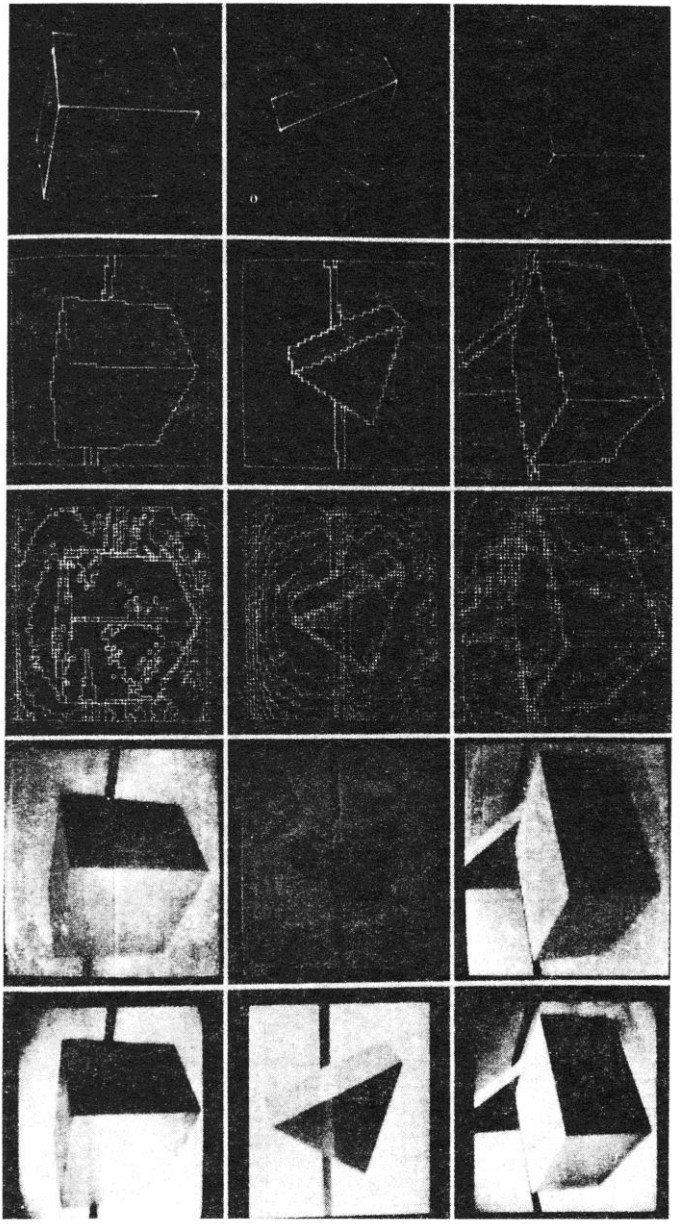


Figura 4.3

en su teoría. La vacuidad de estas «soluciones« es aparente; pero, ¿cómo *podría* ser una solución que no fuera vacía? En un principio puede parecer inconcebible —o por lo menos demasiado complejo como para ser imaginable—, pero éste es precisamente el tipo de obstáculos para la imaginación que debemos afrontar y vencer. Finalmente, resulta más fácil de lo que pudiera parecer el imaginar de qué manera ponía en práctica Shakey sus cualidades sin la ayuda de un *homo ex machina.*

En particular, ¿cómo distinguía Shakey las cajas de las pirámides con la ayuda de su ojo de televisión? La respuesta, a grandes rasgos, era directamente accesible al observador, quien podía presenciar el proceso en la pantalla del ordenador. Una única imagen de televisión, la imagen de una caja, por ejemplo, aparecía en el monitor; la imagen era entonces purificada y rectificada hasta hacerla más nítida y, después, como por arte de magia la silueta de la caja aparecía delineada con unas claras líneas blancas; la imagen se convertía en un dibujo lineal (figura 4.3, página 100).

Acto seguido, Shakey analizaba el dibujo lineal; cada vértice era identificable ya fuese como un vértice de tipo L, de tipo T, de tipo X, de tipo flecha o de tipo Y. Si se descubría un vértice de tipo Y, entonces el objeto tenía que ser una caja, y no podía ser una pirámide; una pirámide no podría proyectar un vértice de tipo Y desde ningún punto de vista.

Esto no deja de ser una simplificación, pero sirve como ilustración de los principios generales sobre los que se basaba el sistema; Shakey tenía un programa de «semántica de líneas» que le permitía manejar las reglas generales para determinar la categoría del objeto cuya imagen aparecía en el monitor. Al mirar el monitor, era de esperar que los observadores sintieran una cierta inquietud al constatar que estaba ocurriendo algo extraño: estaban presenciando un proceso de transformación de imágenes en un monitor, pero *Shakey no estaba mirando al monitor.* Shakey no estaba mirando ni a ese ni a otro monitor en el que las mismas imágenes estuvieran siendo transformadas y analizadas. El equipo sólo disponía de ese monitor y, para el caso, éste podía apagarse o desenchufarse sin que ello afectara al proceso de análisis perceptivo de Shakey. ¿Acaso era este monitor un truco para engañarnos? ¿Para quién había sido ideado? Para los observadores. ¿Qué relación había, pues, entre lo que se podía ver en el monitor y los acontecimientos que se producían en el interior de Shakey?

El monitor era para los observadores, pero la *idea* del monitor era también para los diseñadores de Shakey. Considérese la tarea ingente a la que se enfrentaban: ¿cómo se puede tomar la información de salida de una simple cámara de televisión y extraer de ahí una serie de procesos de identificación de cajas que sean fiables? De los muchos encuadres posibles que la cámara podía enviar al ordenador, sólo un pequeño subconjunto de los mismos correspondía a imágenes de cajas; cada cuadro consistía simplemente en una matriz de celdillas blancas y negras o *píxels*, encendidos y apagados, ceros y unos. ¿Cómo se podía escribir un programa capaz de identificar solamente aquellos cuadros que correspondían a imágenes de cajas? Simplifi-

cando un poco, supóngase que la retina de la cámara fuera una parrilla de 10.000 *píxels*, de 100 por 100. En este caso, cada cuadro sería una de las posibles secuencias de 10.000 ceros y unos. ¿Qué patrones de ceros y unos se formarían entonces ante la presencia de una caja?

En primer lugar, piense en todos esos ceros y unos distribuidos en una matriz, que reproduzca la imagen de la cámara en el espacio, como en la matriz de *píxels* visible en el monitor. Numere los *píxels* en cada fila de izquierda a derecha, como las palabras en una página (y a diferencia de la televisión comercial, cuyo escáner procede en zigzag). Observe, entonces, que las regiones oscuras se componen mayormente de ceros y las regiones claras de unos. Además, la frontera *vertical* entre una región clara a la izquierda y una región oscura a la derecha puede ser descrita de forma bastante simple en términos de una secuencia de ceros y unos: una secuencia formada principalmente por unos hasta el *píxel* número n, seguida por una secuencia formada principalmente de ceros, seguida, exactamente 100 dígitos después (en la siguiente línea), por otra secuencia formada principalmente de unos hasta el *píxel* $n+100$, seguida mayormente por ceros y así, sucesivamente, en múltiplos de 100.

```
0000100000100000100000110111011111101111111011
0010000100000010000000011101011111011111110110111
0100000001010000000100111110101110101111111101
0000010000010000000000011010111111111111101111110
0100000100000001000000001101011111011111111111011
0000000000001000000000001111111011111111111011111
0000000010000000000000011110111111111111111111111
0000000000000001000000011111111110111111111111111
0000000010000000000000101111111111111111011111111
0000100000000000000000011111111111111101111111110
0000000000000000000001001111101111111111111111111
0000001000000000000000011111111111101111111111111
```

Figura 4.4

```
0000000000000000000000110000000000000000000000
0000000000000000000000110000000000000000000000
0000000000000000000000110000000000000000000000
0000000000000000000000110000000000000000000000
0000000000000000000000110000000000000000000000
0000000000000000000000110000000000000000000000
0000000000000000000000110000000000000000000000
0000000000000000000000110000000000000000000000
0000000000000000000000110000000000000000000000
0000000000000000000000110000000000000000000000
0000000000000000000000110000000000000000000000
0000000000000000000000110000000000000000000000
```

Figura 4.5

Un programa capaz de capturar tales periodicidades en el flujo de dígitos proveniente de la cámara podría localizar las fronteras verticales. Una vez localizada, la frontera puede convertirse en una línea vertical blanca y nítida a través de un prudente proceso de sustitución de ceros por unos y viceversa, de manera que una secuencia como 00011000 se repita exactamente cada cien posiciones en la secuencia completa.

Una frontera de claro/oscuro horizontal es igualmente fácil de detectar: un lugar en la secuencia donde una vorágine de unos se ve replicada 100, 200 y 300 dígitos mas adelante (etc.) por una vorágine de ceros.

```
000100000000100000000000100000000010000000000
000000000010100000000000000010000000000000000001
000000001000000000001000000000000001000000000010
000010000000000001000000000000000010000000100000
000000000100000000000000010000000000000000000000
000000000000000000010000000001000000000000000000
111111111101111111111111111011111111111111001111
111111111111111111011111111111111111011111111111
111111111111111111111111110111111111111111111101111
101111111111111111111111111111111111111111111111111
111111111111111111111111101111111111111111111111111
111111111101111111111111111111111111111101111111111
```

Figura 4.6

Las fronteras en pendiente son un poco más difíciles; el programa debe buscar una progresión en la secuencia. Una vez todas las fronteras han sido localizadas y dibujadas en blanco, el dibujo lineal está completo, y se pone en marcha el paso siguiente, y también el más complejo: se «colocan» «plantillas» sobre segmentos de las líneas a fin de identificar los vértices. Una vez hecho esto, utilizar el programa de semántica de líneas para categorizar el objeto que aparece en la imagen es una tarea bastante sencilla; tan sencilla a veces como buscar un único vértice de tipo Y.

Hay numerosos aspectos de este proceso que son relevantes para nuestros propósitos. En primer lugar, cada uno de los subprocesos es «tonto» y mecánico. Es decir, ninguna parte del ordenador tiene que comprender lo que está haciendo y por qué, y, además, no hay misterio alguno sobre cómo se lleva a cabo mecánicamente cada uno de los pasos del proceso. No obstante, una organización inteligente de esos procesos tontos y mecánicos produce un dispositivo que *ocupa el lugar de* un observador entendido. (Colóquese todo el sistema de visión en una «caja negra» cuya función es la de «decir a Shakey lo que necesita saber» sobre lo que tiene enfrente de sí, basado en encuadres de TV que funcionan como información de entrada. Al principio nos sentiremos inclinados a pensar que la única manera de hacerlo es poner un hombrecillo dentro de la caja negra, que mire la pantalla. Ahora vemos que este homúnculo, con sus limitadas tareas, puede ser sustituido por una máquina.)

Una vez vemos cómo se hace, también podemos ver que mientras el proceso es fuertemente *análogo* a un proceso en el que realmente se miran (y se dibujan y se borran) puntos blancos y negros en una pantalla, la localización real en el ordenador de cada una de las operaciones de cambio de ceros en unos y viceversa no es importante, siempre y cuando los números que funcionan como «direcciones» temporales de los dígitos individuales codifiquen la información de qué *píxel* está junto a qué *píxel*. Supóngase que apagamos el monitor. En este caso, pese a que no hay (o no tiene por qué haber) una imagen bidimensional real localizable en el espacio interior del ordenador (en tanto que, pongamos por caso, «patrón de excitación en el *hardware*»), las operaciones son homomórficas (paralelas) a los acontecimientos que observábamos en el monitor. Esos acontecimientos son los únicos que contienen realmente imágenes: una superficie bidimensional de puntos fosforescentes excitados dibujando una forma de tamaño, color, localización y orientación determinados. Así pues, en *un* sentido estricto, Shakey *no* detecta cajas a través de una serie de transformaciones de imágenes; la última imagen real en el proceso es la que aparece dentro del campo receptivo de la cámara. En otro sentido estricto, pero metafórico, Shakey *sí* que detecta cajas a través de una serie de transformaciones de imágenes: el proceso que acabamos de describir, que convierte fronteras de claro/oscuro en un dibujo lineal y después categoriza los vértices. El hecho de que este sentido estricto es, a pesar de todo, metafórico puede evidenciarse al observar que existe una serie de propiedades que sería de esperar que las imágenes *reales* poseyeran y de las que las «imágenes» transformadas por Shakey carecen: no tienen color, ni tamaño, ni orientación. (Con esta imagen podríamos plantear una bonita adivinanza: pienso en una imagen que no es ni mayor ni menor que la *Mona Lisa*, que no es ni en color ni en blanco y negro, y que no se orienta hacia ninguna dirección de la brújula. ¿Qué es?)

El proceso que Shakey utilizaba para extraer información acerca de los objetos de su entorno a partir de la luz apenas se parece al proceso de la visión humana, ni probablemente al proceso visual de cualquier otra criatura. Podemos ignorar este punto por el momento, sin embargo, con el fin de contemplar una posibilidad bastante abstracta en cuanto al modo en que pueden ser descubiertas en el cerebro las imágenes mentales de que nos informan los sujetos humanos. La exposición que hicimos del sistema de visión de Shakey ha sido deliberadamente simplificada a fin de subrayar los aspectos teóricos básicos. Ahora vamos a hacer un poco de ciencia-ficción para establecer un nuevo punto: supóngase que cruzamos a Shakey con otro personaje famoso de la inteligencia artificial, el sistema SHRDLU de Terry Winograd (1972), el cual manipulaba bloques (imaginarios) y después respondía preguntas sobre lo que estaba haciendo y por qué. La mayoría de las respuestas de SHRDLU estaban «enlatadas»: frases o patrones-frase prefabricadas por el propio Winograd. El objetivo de SHRDLU era explorar de forma abstracta alguna de las tareas de tratamiento de la información a que se enfrenta cualquier interlocutor, no elaborar un modelo realista de la pro-

ducción del habla humana, lo cual se aproxima bastante al espíritu de nuestro experimento mental. (En el capítulo 8 examinaremos modelos de producción del habla más realistas.) Un intercambio con nuestra nueva versión de Shakey, rediseñado para manejar un repertorio más sofisticado de acciones verbales, podría proceder de la manera siguiente:

¿Por qué has movido la rampa?
PARA SUBIR POR LA PLATAFORMA.
¿Y por qué querías hacer esto?
PARA EMPUJAR LA CAJA.
¿Y por qué querías hacer esto?
PORQUE TÚ ME LO HAS PEDIDO.

Pero supóngase que ahora preguntáramos a Shakey:

¿Cómo distingues las cajas de las pirámides?

¿Qué frases deberíamos diseñar para que Shakey nos «respondiera»? He aquí tres posibilidades:

1. Examino las secuencias de 10.000 dígitos compuestas de ceros y unos que envía mi cámara, busco patrones en las secuencias, como por ejemplo (...) blablablá (una respuesta muy larga si queremos que Shakey entre en detalles).
2. Busco fronteras de claro/oscuro y en el ojo de mi mente dibujo líneas blancas para marcarlas; después examino los vértices; si, por ejemplo, encuentro un vértice de tipo Y, sé que he dado con una caja.
3. No lo sé; hay cosas que se me antojan como cajas. Simplemente me ocurre. Lo hago por *intuición*.

¿Cuál es la respuesta más adecuada que debería dar Shakey? A su manera, las tres respuestas son ciertas; son descripciones del procesamiento de la información a niveles diferentes, con diferentes grados de precisión o profundidad. El tipo de respuesta que consideremos apropiado introducir en Shakey es en gran medida una cuestión de diseño relacionada con el grado de acceso a los procesos perceptivos con que queremos dotar a la capacidad expresiva de Shakey (la caja negra SHRDLU). Puede que haya buenos motivos de ingeniería para negar un acceso profundo (detallado, costoso en términos de tiempo) a los procesos intermedios de análisis. Sin embargo, cualesquiera que sean las capacidades autodescriptivas de que dotemos a Shakey, habrá un límite en cuanto a la profundidad y el detalle de los «conocimientos» sobre lo que está ocurriendo en su interior y que le será posible expresar. Si la mejor respuesta que puede dar es la 3, entonces se encuentra en la misma posición en cuanto al problema de la discriminación de pirámides y cajas en que nos encontramos nosotros cuando se nos pregunta cómo dis-

tinguimos la palabra «sol» de la palabra «son»; no sabemos cómo lo hacemos; una suena como «sol» y la otra como «son» —más no podemos decir—. Y si se diseña a Shakey para dar respuestas como la de 2, habrá otras preguntas a las que no podrá responder, como por ejemplo «¿cómo dibujas líneas blancas en tus imágenes mentales?». O «¿cómo identificas un vértice de tipo flecha?».

Supóngase que diseñamos a Shakey para tener un acceso de tipo 2 a sus procesos de análisis perceptivo; cuando le preguntemos sobre ello, nos hablará sobre las transformaciones de imágenes que lleva a cabo. Sin que él lo sepa, desenchufamos el monitor. ¿Tenemos entonces derecho a decirle que nosotros sabemos más que él? ¿Él no está procesando imágenes, aunque piense que sí? (*Dice* que lo hace, así que, siguiendo la estrategia heterofenomenológica, interpretaremos esto como una expresión de sus creencias.) Si fuera la simulación realista de una persona, podría replicar que no teníamos ningún derecho a decirle *a él* lo que estaba ocurriendo en su propia mente. *¡Él* sabe lo que está haciendo, lo que está haciendo *realmente!* Si fuera un poco más refinado, podría convenir en que lo que estaba haciendo sólo se podía describir de forma alegórica como procesamiento de imágenes, pese a tener una fuerte tendencia a describirlo de esa manera. En este caso, podríamos decirle que su manera metáforica de presentarlo era bastante atinada.

Si, por otra parte, fuéramos un poco más diabólicos, podríamos trucar a Shakey para que llevara a cabo una charla totalmente espúrea sobre lo que estaba haciendo. Podríamos diseñarlo para querer decir cosas sobre lo que estaba ocurriendo en su interior que no tuvieran la más mínima relación con lo que realmente ocurría. («Utilizo la información de mi TV para dirigir mi cincel interno, el cual modela una forma tridimensional a partir de un bloque de arcilla mental. Entonces, si mi homúnculo se puede sentar encima, es una caja; si se cae, es una pirámide.») No existe ninguna interpretación que conserve el valor de verdad de un testimonio como éste; Shakey estaría sólo *fabulando*, construyendo una historia sin «darse cuenta».

Y en nuestro caso, esta posibilidad demuestra por qué debemos tomarnos la molestia de dar el rodeo de considerar la heterofenomenología como si se tratara de la interpretación de una ficción. Como ya hemos podido comprobar, hay ocasiones en que las personas simplemente se equivocan al valorar lo que están haciendo y cómo lo están haciendo. Y no es que *mientan* ante la situación experimental, sino que se dedican a fabular; llenan huecos, conjeturan, especulan, confunden teorizar con observar. La relación entre lo que dicen y aquello que les hace decir lo que dicen no podría ser más oscura tanto para nosotros, los heterofenomenólogos que *lo vemos desde fuera*, como para los sujetos mismos. *Ellos* no tienen ninguna manera de «ver» (presumiblemente, con un ojo interior) los procesos que gobiernan sus aseveraciones, lo cual no les impide tener sinceras opiniones que expresar.

Resumiendo, los sujetos son creadores inconscientes de ficciones, pero decir que son inconscientes viene a ser el reconocimiento de que lo que di-

cen es, o puede ser, un reflejo *exacto de cómo les parece a ellos que es*. Nos cuentan *cómo es* para ellos resolver el problema, tomar la decisión, reconocer el objeto. Porque son (aparentemente) sinceros, aceptamos que así debe ser para ellos, de lo que se deduce que ello no puede ser más que una incierta guía de lo que está ocurriendo en su interior. A veces es posible demostrar que a pesar de todo las ficciones inconscientes que creamos como sujetos son verdaderas, si dejamos espacio para una cierta relajación metafórica como hicimos en el caso de la respuesta de tipo 2 de Shakey. Por ejemplo, recientes investigaciones en el marco de la psicología cognitiva sobre el problema de las imágenes han demostrado que nuestros testimonios introspectivos sobre las imágenes mentales que experimentamos (sean vacas de color violeta o pirámides) no son totalmente falsos (Shepard y Cooper, 1982; Kosslyn, 1980; Kosslyn, Holtzman, Gazzaniga y Farah, 1985). Trataremos este asunto con mayor detalle en el capítulo 10, donde veremos de qué modo *pueden* interpretarse nuestros testimonios sobre la imaginería mental para que éstos se hagan verdaderos. Sin embargo, como en el caso del Fenhomo terrenal, que resulta no ser capaz de volar ni poseer el don de la ubicuidad, las cosas que encontremos en el cerebro y que *identifiquemos* como imágenes mentales no tendrán las propiedades maravillosas que con tanta seguridad los sujetos han atribuido a sus imágenes. Las «imágenes» de Shakey constituyen un ejemplo claro de cómo algo que no es una imagen puede ser aquello de lo que estamos hablando como si de una imagen se tratara. Aunque los procesos cerebrales que subyacen a la imaginería humana probablemente no se parezcan mucho a los procesos de Shakey, hemos abierto una vía para todo un abanico de posibilidades que de otra manera resultaba difícil de imaginar.

8. La neutralidad de la heterofenomenología

Al comenzar este capítulo prometíamos describir un método, el método heterofenomenológico, que permanecía neutral ante cualquier discusión que enfrentara a enfoques subjetivos y objetivos de la fenomenología, y ante la realidad física o no física de las entidades fenomenológicas. Hagamos un breve repaso del método a fin de comprobar que es realmente así.

En primer lugar, ¿qué actitud tomar ante el problema de los zombíes? Muy fácil, la heterofenomenología por sí misma no puede distinguir entre zombíes y personas reales y conscientes, y, por tanto, no pretende ni resolver ni descartar el problema de los zombíes. *Ex hypothesi*, los zombíes se comportan exactamente igual que las personas reales y, dado que la heterofenomenología es una manera de interpretar la conducta (incluida la conducta interna de los cerebros, etc.), ésta llegará exactamente al mismo mundo heterofenomenológico sea para Zoe que para Zoe la zombie, su gemela sin conciencia. Los zombíes tienen un mundo heterofenomenológico, lo cual sólo significa que cuando los teóricos lo interpretan, llevan a cabo exacta-

mente la misma tarea, utilizando exactamente los mismos medios, que cuando nosotros interpretamos a nuestros amigos. Evidentemente, como ya hemos señalado, algunos de nuestros amigos podrían ser zombíes. (Me resulta un poco difícil contener la risa ante todo esto, pero ya que algunos de los más serios filósofos se toman muy en serio el problema de los zombíes, no puedo más que hacer otro tanto.)

Seguro que no hay nada malo, ninguna parcialidad, al atribuir un mundo heterofenomenológico a los zombíes, ya que es una concesión realmente pequeña. Éste es el minimalismo metafísico de la heterofenomenología. El método describe un mundo, el mundo heterofenomenológico del sujeto, donde hallamos diversos *objetos* (objetos intencionales, en la jerga filosófica) y en donde les pasan diversas cosas a estos objetos. Si alguien pregunta: ¿qué *son* estos objetos y de qué están *hechos*?, nuestra respuesta *podría* ser: «¡Nada!». ¿De qué está hecho mister Pickwick? De nada. Mister Pickwick es un objeto ficticio, como también lo son los objetos descritos, nombrados y mencionados por el heterofenomenólogo.

«Pero, como teórico, ¿no resulta un tanto embarazoso admitir que se está hablando de entidades ficticias, de cosas que no existen?» En absoluto. Los que se ocupan de la teoría literaria llevan a cabo una tarea intelectual honesta y valiosa describiendo entidades ficticias, como también lo hacen los antropólogos, que estudian los dioses y los brujos de las diferentes culturas. También lo hacen los físicos, a quienes si se les pregunta de qué está hecho un centro de gravedad, responderían, «¡De nada!». Los objetos heterofenomenológicos, como los centros de gravedad o el Ecuador, son *abstracta*, no *concreta* (Dennett, 1987a, 1991a). No son fútiles fantasías, sino trabajosas ficciones teóricas. Además, a diferencia del caso de los centros de gravedad, queda abierta la vía de sustituirlas por *concreta* si los progresos de la ciencia empírica así lo permiten.

Hay dos maneras de estudiar el diluvio universal: uno puede asumir que es pura mitología, pero mitología digna de ser estudiada, o se puede intentar averiguar si detrás de eso efectivamente se oculta una catástrofe meteorológica o geológica. Ambas investigaciones pueden ser científicas, pero la primera es menos especulativa. Si lo que usted quiere hacer es especular sobre la segunda posibilidad, la primera cosa que deberá hacer es llevar a cabo una cuidadosa investigación en la línea de la primera posibilidad a fin de descubrir qué indicios hay. Similarmente, si usted quiere estudiar de qué modo (o incluso si *es cierto que*) los elementos fenomenológicos son eventos en el cerebro, la primera cosa que deberá hacer es elaborar un completo catálogo heterofenomenológico de los objetos. Así se corre el riesgo de ofender a los sujetos (de la misma manera que los antropólogos que estudian a Fenhomo corren el riesgo de ofender a sus informantes), pero es el único modo de evitar esa batalla de «intuiciones» que, en su defecto, pasa por ser fenomenología.

Y ¿qué decir de la objeción de que la heterofenomenología, al partir del punto de vista de la tercera persona, no se ocupa de los problemas *reales*

de la conciencia? Nagel, como vimos, insiste en este punto, como también lo hace el filósofo John Searle, quien explícitamente nos pone sobre aviso en contra de mi enfoque: «Recuerden», nos previene Searle, «en estas discusiones, insistan siempre en el punto de vista de la primera persona. El primer paso en el juego de manos operacionalista se produce cuando intentamos figurarnos de qué manera podemos llegar a *saber* cómo es para los demás» (Searle, 1980, pág. 451). Pero no es esto lo que ocurre. Observe que cuando usted está entre las garras del heterofenomenólogo, *es usted quien tiene la última palabra.* Usted puede corregir, revisar y desdecirse *ad libitum*, y mientras evite *teorizar* presuntuosamente sobre las causas o el estatuto metafísico de los elementos sobre los que usted habla, cualquier cosa en la que insista recibirá autoridad constitutiva para determinar lo que ocurre en su mundo heterofenomenológico. Usted es el novelista, y lo que usted dice es lo que vale. ¿Qué más se puede pedir?

Si usted quiere que *creamos* todo lo que usted dice sobre su fenomenología, entonces no sólo está pidiendo que se le tome en serio, sino que está solicitando infalibilidad papal, y eso ya es pedir demasiado. Usted *no* tiene autoridad sobre lo que está ocurriendo en su interior, sino sobre lo que *parece* estar ocurriendo, y se le concede una autoridad total, dictatorial, para el análisis de cómo le parecen a usted las cosas, para analizar *cómo es para usted.* Y si se queja usted porque algunas partes de lo que le parece son inefables, nosotros los heterofenomenólogos también aceptaremos eso. ¿Qué mejores razones podemos tener para creer que usted no puede describir algo que (1) el hecho de que usted no lo describa, y (2) el hecho de que usted confiese no poder hacerlo? Es evidente que usted puede estar mintiendo, pero le concederemos el beneficio de la duda. Si usted replica, «no sólo estoy diciendo que yo no puedo describirlo; ¡estoy diciendo que no se puede describir!», nosotros los heterofenomenólogos nos limitaremos a tomar nota de que usted no puede describirlo *ahora*, y, puesto que usted es el único en posición de describirlo, en este momento no se puede describir. Puede que más adelante usted sea capaz de describirlo, y en ese momento *será* algo totalmente distinto, será algo que se puede describir.

Cuando yo anuncio que los objetos de la heterofenomenología son las ficciones de un teórico, usted puede tener la tentación (muchos la tienen, por lo que he visto) de contraatacar diciendo:

> Esto es *únicamente* lo que distingue a los objetos de la fenomenología real de los objetos de la heterofenomenología. Mis objetos *auto*fenomenológicos no son objetos ficticios —son perfectamente *reales*—, aunque no tengo ni la menor idea de qué están hechos. Cuando le digo, sinceramente, que estoy imaginando una vaca de color violeta, no estoy sólo produciendo inconscientemente una cadena de palabras a tal efecto (como Shakey), hábilmente ideada para coincidir con algún vago evento físico, análogo en mi cerebro; estoy haciendo una relación, consciente y deliberada, de la existencia de algo que está *realmente ahí.* ¡Para mí no es la mera ficción de un teórico!

Reflexione con detenimiento sobre este discurso. Usted dice que no sólo está produciendo inconscientemente una cadena de palabras. Bueno, pero está produciendo inconscientemente una cadena de palabras; y no tiene ni la más remota idea de cómo lo hace, ni de qué factores intervienen en su producción. Sin embargo, usted insiste en decir que no está haciendo *solamente* eso; usted sabe *por qué* lo hace; usted *comprende* las palabras, y *quiere decir* lo que dice. Estoy de acuerdo. Ésta es la razón por la cual sus palabras funcionan tan bien para construir un mundo heterofenomenológico. Si usted se dedicara a lanzar palabras más o menos al azar, como un loro, las probabilidades en contra de que surgiera una secuencia de palabras con la interpretación deseada serían astronómicas. Seguro que hay una buena explicación de cómo y por qué usted dice lo que dice, una explicación que dé cuenta de la diferencia que hay entre sólo decir algo y decir algo y quererlo decir, *pero usted todavía no tiene esa explicación.* O por lo menos no en su totalidad. (En el capítulo 8 nos ocuparemos de este asunto.) Es probable que usted esté hablando de algo real, al menos la mayor parte del tiempo. Vamos a ver si podemos descubrir qué es.

Estas palabras de aliento no son suficientes para algunas personas. Algunas personas no jugarían con estas reglas. Algunas personas de devota religiosidad, por ejemplo, se ofenden cuando sus interlocutores apenas insinúan que *podría* haber una religión verdadera alternativa. Estas personas no ven neutralidad en el agnosticismo, ven en ello una afrenta, porque uno de los dogmas de su credo considera pecado el no creer; se sienten autorizados (si es que ésta es la palabra correcta) por sus sentimientos heridos a sufrir cuando se encuentran con escépticos o agnósticos, y, a menos que puedan controlar la ansiedad que sienten cuando ven que alguien (todavía) no cree en lo que dicen, se autoexcluyen de la investigación científica académica.

En este capítulo hemos desarrollado un método *neutral* para investigar y describir la fenomenología. Consiste en extraer y depurar *textos* a partir de *sujetos* (aparentemente) parlantes, y en utilizar esos textos para generar una ficción teórica, el *mundo heterofenomenológico* del sujeto. En este mundo ficticio moran todas las imágenes, eventos, sonidos, olores, intuiciones, presentimientos y sensaciones que el sujeto (aparentemente) sinceramente cree que existen en su flujo de la conciencia. Si lo ampliamos al máximo, es un retrato neutral y exacto de *cómo es ser* ese sujeto, en los términos del propio sujeto y dada la mejor interpretación que podemos elaborar.

Una vez han extraído esta heterofenomenología, los teóricos pueden concentrarse en el problema de buscar una explicación para la *existencia* de esta heterofenomenología en todos sus detalles. La heterofenomenología existe —tan claro como que las novelas y otras ficciones existen—. No cabe duda de que las personas creen poseer imágenes mentales, dolores, experiencias perceptivas y todo lo demás, y *estos* hechos —los hechos en que estas personas creen y que relatan cuando expresan sus creencias— son fenómenos de los cuales toda teoría científica de la mente debe dar cuenta. Organizamos nuestros datos en relación a estos fenómenos en forma de ficciones teóri-

cas, «objetos intencionales» en mundos heterofenomenológicos. Así pues, la cuestión de si elementos así descritos existen en tanto que objetos reales, eventos y estados en el cerebro —o, para el caso, en el alma— es susceptible de ser investigada empíricamente. Si damos con los candidatos reales adecuados, podremos identificarlos con los referentes que tanto hemos buscado de los términos empleados por el sujeto; si no, tendremos que explicar por qué a los sujetos les parece que estas entidades existen.

Ahora que hemos puesto en orden nuestros supuestos metodológicos, podemos finalmente concentrarnos en la teoría empírica de la conciencia. Empezaremos enfrentándonos al problema de la ordenación espacial y cronológica de los elementos en nuestro flujo de la conciencia. En el capítulo 5 presentamos un primer esbozo de la teoría y mostramos cómo puede dar cuenta de un caso simple. En el capítulo 6 veremos cómo la teoría nos permite reinterpretar algunos fenómenos mucho más complejos que han suscitado la perplejidad de los teóricos. Del capítulo 7 al 9 desarrollamos la teoría más allá de su primer esbozo, intentando disipar malas interpretaciones y objeciones, y haciendo nuevo hincapié sobre sus virtudes.

SEGUNDA PARTE

UNA TEORÍA EMPÍRICA DE LA MENTE

Capítulo 5

VERSIONES MÚLTIPLES FRENTE AL TEATRO CARTESIANO

1. El punto de vista del observador

> No existe célula o grupo de células en el cerebro cuya preeminencia anatómica o funcional las haga aparecer como la piedra angular o el centro de gravedad de todo el sistema.
>
> WILLIAM JAMES, *Principles of Psychology*, 1890

Los que navegan en barcos de recreo bordeando una costa peligrosa suelen guardarse de sus peligros poniendo proa hacia alguna marca visible. Buscan una boya distante aproximadamente en la dirección en que quieren ir, consultan en la carta marina que no haya algún obstáculo oculto a lo largo de la línea recta que separa la boya del punto en que se encuentran y navegan directos hacia ella. Durante una hora o más, el objetivo del patrón se reduce a navegar en dirección a la marca, corrigiendo cualquier pequeño error en el rumbo. Con frecuencia ocurre, sin embargo, que los patrones se concentran tanto en su proyecto que se olvidan de virar en el último momento ¡y acaban por chocar con la boya! Se distraen del objetivo principal de mantener la embarcación lejos de todo peligro a causa de la tranquilidad que les produce la seguridad de completar con éxito el objetivo más limitado de alcanzar la marca. En este capítulo demostraremos que algunas de las más sorprendentes paradojas de la conciencia son fruto de aferrarse demasiado a un buen hábito de pensamiento que nos mantiene alejados de los escollos.

Siempre que hay una mente consciente, hay un *punto de vista*. Ésta es una de las ideas fundamentales que tenemos sobre la mente —o sobre la conciencia—. Una mente consciente es un observador que recoge un subconjunto limitado de toda la información disponible. Un observador recoge la información que está disponible en una determinada secuencia (aproximadamente) continua de tiempos y lugares en el universo. En la práctica, podemos considerar que el punto de vista de un sujeto consciente determinado es exactamente esto: un *punto* que se mueve en el espacio-tiempo. Considérense, por ejemplo, los diagramas típicos de la física y la cosmología que ilustran el efecto Doppler o los efectos de desviación de la luz a causa de la gravedad.

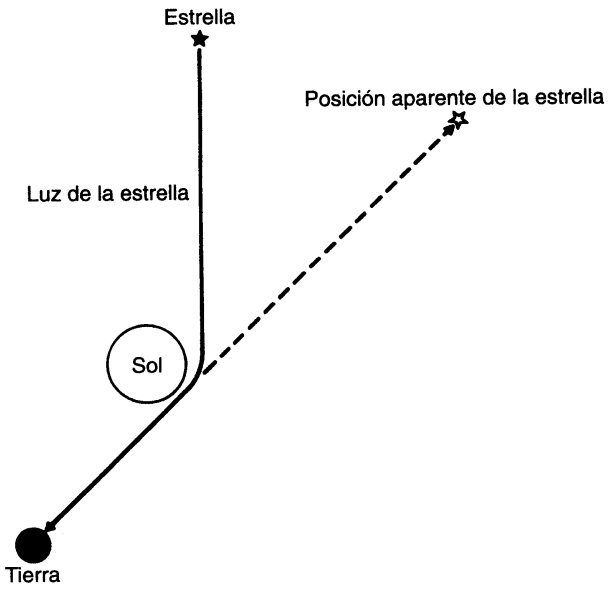


Figura 5.1

En la figura 1, el observador es un punto fijo sobre la superficie de la Tierra. Para observadores en distintos puntos del universo las cosas se verían de otra manera. Existen ejemplos más simples y más conocidos. Por ejemplo, explicamos el tiempo que separa el momento en que vemos unos fuegos artificiales del momento en que los oímos por las diferentes velocidades a que viajan la luz y el sonido. Llegan *hasta el observador* (en el punto que éste ocupa) en momentos diferentes a pesar de haber abandonado el lugar de origen al mismo tiempo.

¿Qué ocurre, sin embargo, cuando nos concentramos en el observador e intentamos localizar su punto de vista de manera más precisa, como un punto *dentro* del individuo? Aquellos supuestos que tan bien funcionan a gran escala empiezan a fallar.[1] No existe un sólo punto en el cerebro al cual

1. Este problema tiene una cierta semejanza con las dificultades que encuentran los físicos cuando se enfrentan a una *singularidad*, un punto en el que, precisamente por carecer de dimensiones, muchas magnitudes adoptan valores infinitos (dada su definición). Esto es lo que pasa con los agujeros negros, pero también afecta a la interpretación de entidades más mundanas. Roger Penrose comenta el caso de cómo aplicar las ecuaciones de Lorentz y las ecuaciones de Maxwell a las partículas. «Lo que las ecuaciones de Lorentz nos dicen que debemos hacer es examinar el campo electromagnético en el *punto* exacto en el que está localizada la partícula (y, en efecto, nos proporcionan una fuerza en dicho punto). ¿Qué punto debe tomarse si la partícula tiene un tamaño finito? ¿Debemos tomar el centro de la partícula, o debemos promediar el campo (de la fuerza) sobre todos los puntos de la superficie? (...) Quizá sea mejor considerar la par-

acuda toda la información, un hecho que tiene algunas consecuencias que distan mucho de ser evidentes —resultan, de hecho, bastante antiintuitivas.

Dado que trataremos con eventos que se producen a escalas relativamente microscópicas de espacio y tiempo, conviene tener claro el tipo de magnitudes que manejaremos. Todos los experimentos que consideraremos comportan intervalos de tiempo medidos en milisegundos o milésimas de segundo. Le será de ayuda tener una idea aproximada de lo largo (o corto) que son 100 mseg o 50 mseg. Usted puede proferir alrededor de cuatro o cinco sílabas por segundo, así que cada sílaba tiene una duración del orden de los 200 mseg. En el cine se proyectan imágenes a una velocidad de veinticuatro fotogramas por segundo, de modo que la película avanza un fotograma cada 42 mseg (de hecho, cada fotograma permanece estacionario y se expone tres veces durante esos 42 mseg, con exposiciones de 8,5 mseg e intervalos de oscuridad de 5,4 mseg entre cada exposición). La televisión (en los EE.UU.) emite a treinta y tres imágenes por segundo, o una imagen cada 33 mseg (en realidad, cada imagen se entreteje en dos pases, de manera que la segunda se solapa con la primera). Moviendo el pulgar tan rápido como le sea posible, se puede poner en marcha y parar un cronómetro en unos 175 mseg. Cuando usted se pega un martillazo en el dedo, las fibras nerviosas rápidas (mielinizadas) envían el mensaje al cerebro en unos 20 mseg; las lentas, las fibras-C no mielinizadas, envían señales de dolor que tardan más tiempo —alrededor de los 500 mseg— en recorrer la misma distancia.

En la siguiente tabla tenemos los valores aproximados en milisegundos de la duración de algunos eventos relevantes:

decir «uno, Mississippi»	1.000 mseg
fibra no mielinizada, del dedo al cerebro	500 mseg
una bola rápida a 145 km/h en recorrer los 18,5 m hasta la posición del bateador	458 mseg
proferir una sílaba	200 mseg
poner en marcha y parar un cronómetro	175 mseg
fotograma de cine	42 mseg
fotograma de televisión	33 mseg
fibra rápida (mielinizada), del dedo al cerebro	20 mseg
el ciclo básico de una neurona	10 mseg
el ciclo básico de un ordenador personal	0,0001 mseg

Descartes, uno de los primeros en pensar seriamente sobre lo que debe ocurrir cuando observamos más de cerca dentro del cuerpo del observador, desarrolló una idea que es tan superficialmente natural y atractiva que ha

---

tícula como un *punto* determinado. Pero esto lleva a otro tipo de problemas, pues entonces el propio campo eléctrico de la partícula se hace *infinito* en su inmediata vecindad.» (Penrose, 1989, págs. 189-190 [245-246 de la edición española; N. del T.])

influido en nuestra manera de pensar sobre la conciencia desde entonces. Como vimos en el capítulo 2, Descartes decidió que el cerebro *tiene* un centro: la glándula pineal, que sirve de pórtico para la mente consciente (véase la figura 2.1, página 45). La glándula pineal es el único órgano del cerebro que se encuentra en la línea media, sin tener una pareja, uno en cada hemisferio. Es la que aparece marcada con una «L» en este diagrama del gran anatomista del siglo XVI, Vesalius. Algo más pequeña que un guisante, está ahí, totalmente aislada sobre su base, conectada al resto del sistema nervioso, justo en el medio de la parte posterior del cerebro. Ya que su función era prácticamente insondable (aun hoy no se sabe muy bien qué hace la glándula pineal), Descartes sugirió una: a fin de que una persona sea consciente de algo, el tráfico desde los sentidos debía llegar a esta estación, donde se llevaba a cabo una transacción especial —mágica, de hecho— entre el cerebro material de la persona y su mente inmaterial.

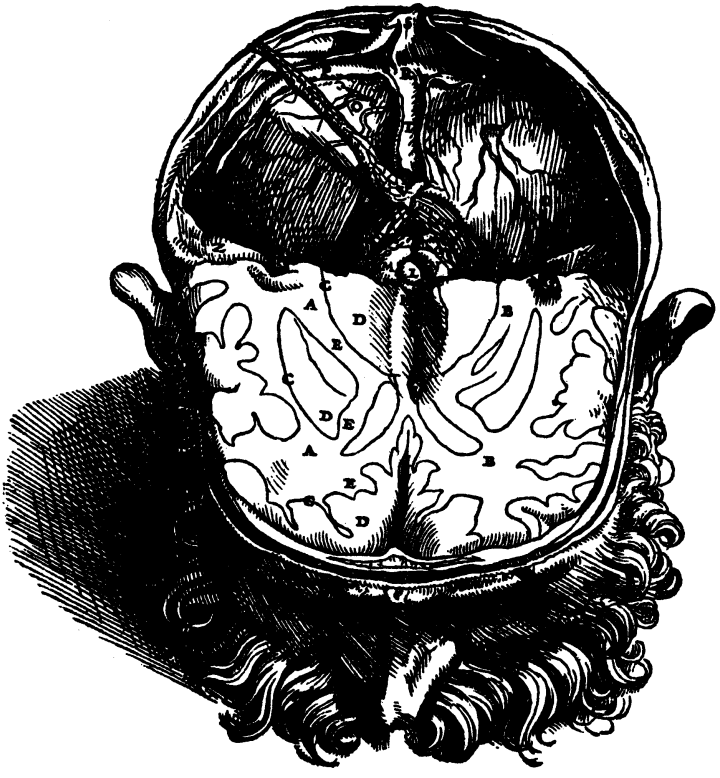


Figura 5.2

Según Descartes, no todas las reacciones corporales requerían esta intervención de la mente consciente. Descartes sabía de la existencia de lo que hoy llamamos reflejos, y propuso que éstos se producían de manera enteramente mecánica, mediante una especie de cortocircuitos que evitaban pasar por la estación pineal, y que eran, por tanto, inconscientes.

Figura 5.3

Estaba equivocado en cuanto a los detalles: pensaba que el fuego desplazaba la piel, lo que provocaba el tirón de un hilo minúsculo, lo que a su vez abría un poro en el ventrículo (F) de donde fluía un «espíritu animal» que, viajando por un tubito, hinchaba los músculos de la pierna y hacía que se retirara el pie (Descartes, 1664). Por lo demás, era una buena idea. No se puede decir lo mismo de la visión de Descartes sobre el papel de la glándula pineal como puerta giratoria de la conciencia (lo que podríamos llamar el cuello de botella cartesiano). Esta idea, el dualismo cartesiano, es totalmente errónea, como vimos en el capítulo 2. Sin embargo, aunque el materialismo de uno u otro tipo es hoy en día la opinión compartida por casi todo el mundo, incluso los materialistas más acérrimos olvidan que una vez la *res cogitans* cartesiana ha sido descartada, ya no hay lugar para un pórtico centralizado

o, en general, para ningún centro *funcional* en el cerebro. La glándula pineal no sólo no es el aparato de fax hacia el alma, sino que tampoco es el despacho oval del cerebro, ni lo son ninguna de las otras partes del cerebro. El cerebro es el cuartel general, allí donde está el último observador, pero no hay ninguna razón para creer que el cerebro posea otro cuartel general más profundo, un santuario interior, el paso por el cual es condición necesaria y suficiente para la experiencia consciente. En pocas palabras, no hay ningún observador dentro del cerebro.[2]

La luz viaja mucho más deprisa que el sonido, como lo demuestra el ejemplo de los fuegos artificiales, pero ahora sabemos que el cerebro tarda más en procesar los estímulos visuales que los auditivos. Como ha señalado el investigador del cerebro Ernst Pöppel (1985, 1988), gracias a estos desequilibrios, el «horizonte de simultaneidad» se sitúa a una distancia de *unos* diez metros: la luz y el sonido que abandonan un mismo punto a diez metros de los órganos sensoriales del observador producen respuestas neuronales que son «centralmente accesibles» al mismo tiempo. ¿Podemos precisar un poco más? Aquí está el problema. El problema no radica solamente en medir las distancias desde el punto exterior en que se ha producido el evento hasta los órganos sensoriales, o la velocidad de transmisión en los distintos medios, o dar cabida a diferencias individuales. El problema fundamental es el de decidir qué debe ser considerado como la «línea de meta» del cerebro. Pöppel obtuvo sus resultados comparando mediciones comportamentales: tiempos de reacción medios (al presionar un botón) a estímulos visuales y auditivos. Las diferencias varían entre los 30 mseg y los 40 mseg, el tiempo que tarda el sonido en viajar diez metros aproximadamente (el tiempo que tarda la luz en viajar diez metros no es significativamente distinto de cero). Pöppel utilizó una línea de meta periférica —la conducta exterior—, pero nuestra intuición nos dice que la *experiencia* de la luz o el sonido se produce en un punto que se halla *entre* el momento en que las vibraciones alcanzan los órganos sensoriales y el momento en que presionamos el botón como señal de que se ha producido la experiencia. Y además, ello se produce en algún *centro,* en algún lugar del cerebro en las vías excitadas entre el órgano sensorial y el dedo. Parece que si pudiéramos decir *exactamente* dónde,

2. Negar que la cabeza es el cuartel general sería una locura, pero no una locura sin precedentes. Phillippe Pinel relata en 1800 el curioso caso de un hombre que cayó en «un verdadero delirio producto de los terrores de la revolución. El trastorno de la razón en este hombre estaba marcado por un hecho singular: creía haber sido guillotinado y que su cabeza había ido a parar a un montón de cabezas de otras víctimas. Después, al arrepentirse los jueces demasiado tarde de sus crueles actos, ordenaron devolver las cabezas a sus respectivos cuerpos. Sin embargo, a causa de un error, pusieron sobre los hombros de este hombre la cabeza de otro desgraciado. Esta idea de que su cabeza había sido cambiada le ocupaba noche y día... «¡miren mis dientes!, repetía sin cesar, ¡eran perfectos, y ahora están todos podridos! ¡Mi boca estaba sana, pero ésta está infectada! ¡Qué diferencia entre este pelo y el pelo que tenía antes de que me cambiaran la cabeza!» *Traité médico-philosophique sur l'aliénation mentale, ou la Manie.* París: Chez Richard, Caille et Ravier, 1800, págs. 66-67. (Debo agradecer a Dora Weiner el haber puesto en mi conocimiento este caso tan fascinante.)

podríamos decir exactamente cuándo se produjo la experiencia. Y viceversa, si pudiéramos decir exactamente cuándo ocurrió, podríamos decir dónde se localiza la experiencia consciente.

Llamemos a esta idea de que existe un lugar central en el cerebro *materialismo cartesiano*, pues es la visión a que se llega cuando se rechaza el dualismo de Descartes pero no se consigue abandonar esa imagen de un teatro central (aunque material) adonde «todo acude». La glándula pineal podría ser uno de los candidatos a ser este Teatro Cartesiano, pero también se han propuesto otros como el cingulado anterior, la formación reticular o varios puntos en los lóbulos centrales. El materialismo cartesiano es la tesis según la cual existe una línea de meta crucial o una frontera en algún punto del cerebro, señalando el lugar en que el orden de llegada equivale al orden de «presentación» en la experiencia, porque *lo que allí tiene lugar* es aquello de lo que usted es consciente. Es posible que hoy en día ya nadie acepte explícitamente el materialismo cartesiano. Muchos teóricos insistirían en afirmar que han rechazado explícitamente una idea tan mala como ésta. Pero, como veremos, la persuasiva imagen del Teatro Cartesiano sigue volviendo para perseguirnos —tanto a profanos como a científicos— incluso mucho después de haber denunciado y exorcizado al fantasmagórico dualismo.

El Teatro Cartesiano es una manera metafórica de explicar el modo en que la experiencia consciente se localiza en el cerebro. En un principio, parece ser una extrapolación inocente del conocido e innegable hecho de que para *intervalos de tiempo macroscópicos normales*, efectivamente podemos clasificar los acontecimientos en dos categorías: «aún no observado» y «ya observado». Llevamos a cabo esta operación localizando al observador en un punto y trazando las trayectorias de los vehículos de información relativas a este punto. Sin embargo, cuando tratamos de extender este método para explicar fenómenos que se producen en intervalos muy breves de tiempo, nos encontramos con una dificultad *lógica*: si el «punto» de vista del observador debe esparcirse sobre una gran superficie en el cerebro de éste, la propia sensación subjetiva del observador de secuencia y simultaneidad *debe* poder determinarse mediante algo más que el «orden de llegada», ya que el orden de llegada no estará definido completamente hasta que el punto de destino haya podido ser determinado. Si A llega antes que B a una determinada meta, pero B llega antes que A a otra, ¿qué resultado debe tomarse para fijar la secuencia subjetiva en la conciencia? (Véase Minsky, 1985, pág. 61.)* Pöppel habla de los momentos en que la visión y el sonido se hacen «centralmente accesibles» en el cerebro, pero ¿qué punto o puntos de «accesibilidad central» «contarían» para determinar el orden *experimentado*, y por qué? Cuando intentemos responder a esta pregunta, nos veremos obligados a abandonar la idea del Teatro Cartesiano y a sustituirla por un modelo nuevo.

La idea de un centro especial en el cerebro es una de las más tenaces y perniciosas de las que acosan nuestros intentos de pensar sobre la con-

* Página 63 de la edición castellana (Galápago, 1986). [N. del T.]

ciencia. Como veremos, sigue reapareciendo, bajo diversas formas, y por una serie de motivos aparentemente irresistibles. En primer lugar, está nuestra apreciación personal e introspectiva de la «unidad de la conciencia» que nos empuja a establecer esa distinción entre el «aquí dentro» y el «ahí afuera». La ingenua frontera entre el «yo» y «el mundo exterior» es mi piel (y los cristalinos de mis ojos), pero, a medida que adquirimos conciencia de la accesibilidad de los acontecimientos que tienen lugar en nuestros cuerpos, el gran exterior nos invade. «Aquí dentro» puedo intentar levantar mi brazo, pero «ahí afuera» se «me ha dormido» o está paralizado, no se va a mover; mis líneas de comunicación desde dondequiera que *yo* esté hasta la maquinaria neuronal que controla mi brazo han sido intervenidas. Y si mi nervio óptico fuera seccionado, yo no esperaría seguir viendo aunque mis ojos permanecieran intactos; el poseer experiencias visuales es algo que aparentemente se produce *en el interior* de mis ojos, en algún punto entre mis ojos y mi voz cuando le cuento a usted lo que veo.

¿No se deduce entonces, *en virtud de una necesidad geométrica*, que nuestras mentes conscientes se encuentran *al cabo* de todos los procesos *in*ternos, justo antes del *inicio* de los procesos *ex*ternos que realizan nuestras acciones? Avanzando desde la periferia por los canales de entrada de mi ojo, por ejemplo, ascendemos por el nervio óptico y más arriba hasta diversas zonas del córtex visual, y ¿entonces...? Ambos trayectos avanzan el uno hacia el otro por dos planos inclinados, el aferente (la entrada) y el eferente (la salida). Por muy difícil que sea determinar en la práctica la localización precisa de la divisoria continental del cerebro, ¿no debe haber acaso, por pura extrapolación geométrica, un punto máximo, un punto focal, un punto tal que todo lo que se entromete por un lado es *preexperiencial*, y todo lo que se entromete por el otro es *postexperiencial*?

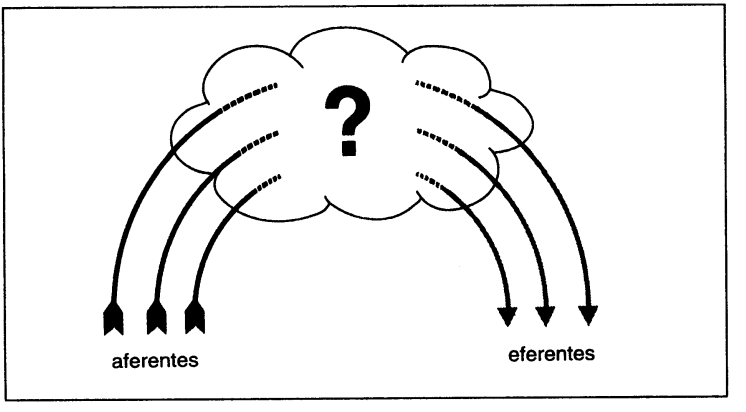


Figura 5.4

En la concepción de Descartes, esto es fácilmente visible, ya que todo fluye hacia y desde la estación pineal. Parecería, pues, que si adoptáramos un modelo más actual del cerebro, deberíamos poder marcar nuestras exploraciones con colores, utilizando, pongamos por caso, el rojo para las vías aferentes y el verde para las vías eferentes; allí donde los colores cambiaran sería el punto medio funcional en la gran divisoria mental.

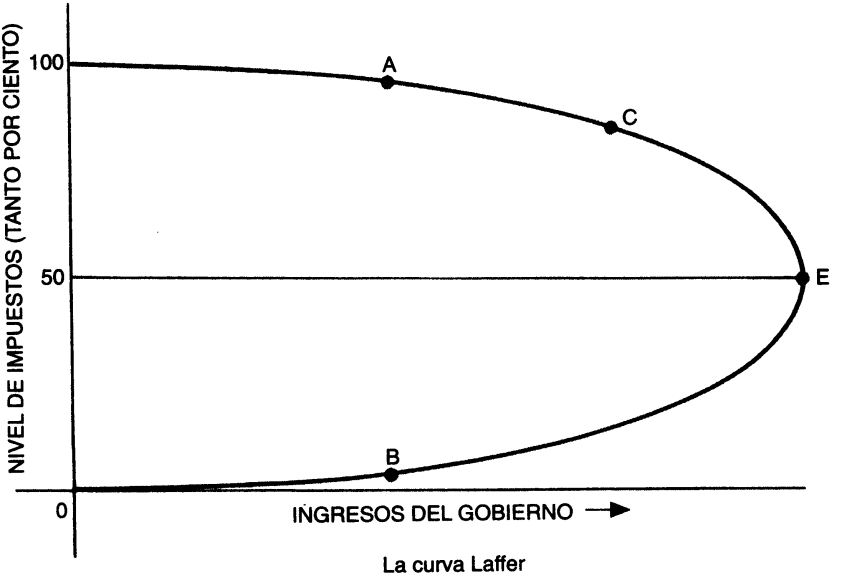
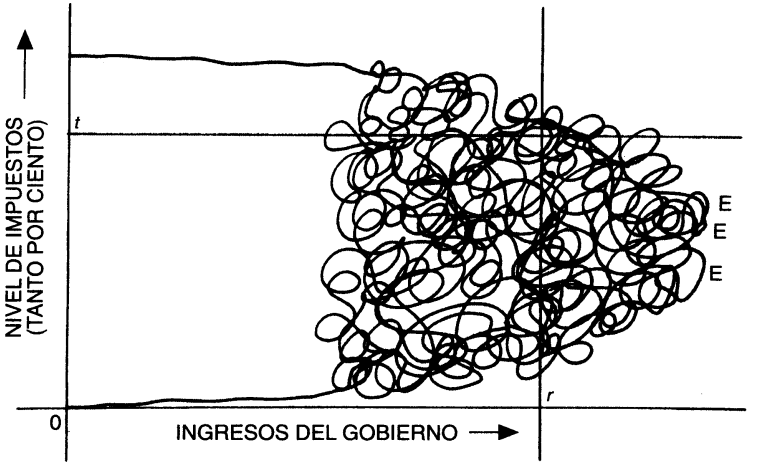


Figura 5.5.

Este argumento, cuyo atractivo no deja de ser curioso, quizás a algunos de ustedes les resulta familiar. Es el hermano gemelo de otro argumento igualmente falaz, que últimamente ha resultado tener un peso excesivo: la conocida curva de Arthur Laffer, el fundamento intelectual (si se me permite faltar al rigor lingüístico) de la Reaganomics. Si el nivel impositivo es 0, el gobierno no obtiene ingresos, y si el nivel impositivo es del 100 %, nadie trabajaría por un salario, y el gobierno tampoco obtendría ingresos; a un 2 % los ingresos del gobierno serían prácticamente el doble que a un 1 % y así sucesivamente, pero a medida que el nivel de impuestos sube, el nivel de ingresos empieza a bajar; los impuestos son onerosos. Si observamos el otro lado de la escala, un 99 % de impuestos daría lugar a una recaudación

muy poco más ventajosa que el 100 %, de modo que los ingresos apenas aumentarían; a un 90 % la situación del gobierno mejoraría, y mejor aún sería a un 80 %. El trazado de ambas partes de la parábola no tiene por qué ser simétrico, pero ¿no debería existir, en virtud de una necesidad geométrica, un punto de inflexión en la curva, un nivel de impuestos que maximice los ingresos? La idea de Laffer era que, dado que en aquel momento el nivel de impuestos estaba en la parte superior de la curva, la bajada de los impuestos provocaría una subida de los ingresos. Era una idea tentadora; a muchos les pareció que debía ser correcta. Pero, como ha señalado Martin Gardner, el mero hecho de que los extremos de la curva sean conocidos no es garantía de que la parte no conocida de la misma en sus regiones centrales siga un curso regular. A modo de sátira, Gardner propone como alternativa la «curva neo-Laffer», que tiene más de un máximo, de modo que la posibilidad de acceder al uno o al otro depende de complejos factores relacionados con la historia y la coyuntura que la alteración de una única variable no sería capaz de determinar (Gardner, 1981). Deberíamos aplicar la misma moraleja al caso de la espesa bruma que nos impide ver adónde van y de dónde vienen las vías periféricas aferentes y eferentes: la claridad de las vías periféricas no nos garantiza que las mismas distinciones se mantengan *siempre en el camino hacia el interior.* La «maraña técnica» que Gardner prevé para la economía no es nada comparada con el revoltijo de actividades que tiene lugar en las regiones centrales del cerebro. Debemos dejar de pensar en el cerebro como si tuviera esa única cumbre funcional o punto



La curva neo-Laffer (NL)

Figura 5.6

central. Porque esto no es un atajo inocuo; es un mal hábito. A fin de acabar con este mal hábito de pensamiento, es preciso que examinemos algunos casos de ese mal hábito en acción, pero también necesitamos una buena imagen que lo sustituya.

2. Presentación del modelo de Versiones Múltiples

He aquí un primer esbozo del sustituto, el modelo de Versiones Múltiples de la conciencia. En un principio, es de esperar que resulte un tanto extraño y difícil de visualizar —no en vano la idea del Teatro Cartesiano está tan arraigada—. De acuerdo con el modelo de las Versiones Múltiples, todas las variedades de la percepción —de hecho, todas las variedades del pensamiento y la actividad mental— se llevan a cabo en el cerebro a través de procesos paralelos, que corren por múltiples vías, de interpretación y elaboración de los estímulos sensoriales de entrada. La información que entra en el sistema nervioso se halla sometida a un continuo proceso parecido al de una compilación editorial. Por ejemplo, ya que su cabeza se mueve poco y sus ojos se mueven mucho, las imágenes en sus retinas se balancean constantemente, un poco como las imágenes de esos vídeos domésticos grabadas por personas con pulso poco firme. Pero no es así como las cosas nos aparecen. Normalmente las personas se sorprenden cuando descubren que, bajo condiciones normales, sus ojos se mueven en rápidas sacudidas, del orden de unas cinco fijaciones rápidas por segundo, y que este movimiento, como el movimiento de sus cabezas, se corrige durante el período de procesamiento que va desde el globo ocular a... la conciencia. Los psicólogos han aprendido mucho sobre los mecanismos que se ocupan de conseguir estos efectos normales, y han descubierto también algunos efectos especiales, como el de la interpretación de la profundidad en estereogramas de puntos aleatorios (Julesz, 1971). (Véase figura 5.7, página 126.)

Si usted mira estos dos cuadrados ligeramente distintos a través de un estereoscopio (o si simplemente los mira con los ojos bizcos hasta conseguir que ambas imágenes se fundan en una sola —algunas personas son capaces de hacerlo sin la ayuda de ningún dispositivo), usted acabará por ver una forma que emerge en tres dimensiones, gracias a un impresionante proceso compilador producido en el cerebro que compara y coteja la información que recibe de cada ojo. Se puede llegar a obtener un registro global óptimo sin necesidad de someter cada matriz de datos a un elaborado proceso de extracción de rasgos. A bajo nivel, existen suficientes coincidencias destacables —cada uno de los puntos en un estereograma de puntos aleatorios— para dictar una solución.

Los procesos compiladores del cerebro necesitan bastante tiempo para producir estos efectos, pero existen otros efectos especiales que son más rápidos. El efecto McGurk (McGurk y Macdonald, 1979) sería un ejemplo de estos últimos. Cuando una película francesa se dobla al inglés, los especta-

Figura 5.7

dores en su mayoría no se dan cuenta de las discrepancias que existen entre los movimientos de labios que ven y los sonidos que oyen (siempre que el doblaje no sea muy malo). Pero, ¿qué ocurre si se crea una banda sonora que encaja perfectamente con las imágenes, con la excepción de algunas consonantes deliberadamente discordantes? (Recuperando un concepto ya utilizado en el capítulo anterior con fines distintos, podemos suponer que sería una versión filmada de los labios de una persona diciendo «de izquierda a derecha» y una banda sonora en que la voz dijera «d'ezquierda a lerecha».) ¿Qué experimentaría la gente? *Oirían* «de izquierda a derecha». En este caso, en la competición compiladora artificialmente inducida entre la información que proviene de los ojos y la que proviene de los oídos, ganarían los ojos.[3]

Estos procesos de edición se producen durante fracciones de segundo importantes, y en ese tiempo se pueden producir, en diversos órdenes, varios añadidos, incorporaciones, enmiendas y sobreescrituras de contenido. No experimentamos directamente lo que ocurre en nuestras retinas, en nuestros oídos, en la superficie de nuestra piel. Lo que realmente experimentamos es el producto de muchos procesos interpretativos —los procesos de compilación. Éstos operan sobre representaciones relativamente burdas y simples, y devuelven representaciones cotejadas, revisadas y amplificadas, y tienen lugar en los flujos de actividad que se producen en distintas partes del cerebro. Hasta aquí coinciden casi todas las teorías de la percepción, pero aquí es donde entran en acción los aspectos novedosos del modelo de Versiones Múltiples: los procesos de detección de rasgos o de discriminación

3. Un ejemplo aún más sorprendente es aquel experimento en que se hace creer al sujeto, mediante un juego de espejos, que es su propia mano la que está dibujando una línea, cuando la verdad es que lo que está viendo es la mano de un ayudante del investigador. En este caso «los ojos ganan» hasta tal punto que el proceso de edición en el cerebro es inducido a concluir que la mano del sujeto está siendo movida a la fuerza; éste afirma sentir la «presión» que le impide mover «su» mano en la dirección deseada (Nielsen, 1963).

tan sólo tienen que efectuarse una vez. Es decir, cuando una porción especializada y localizada del cerebro ha llevado a cabo la «observación» de un rasgo determinado, el contenido informativo queda fijado y no tiene por qué ser enviado a alguna otra parte para ser rediscriminado por un «maestro» discriminador. En otras palabras, el proceso de discriminación no conduce a una *representación* del rasgo discriminado en beneficio de los espectadores del Teatro Cartesiano, porque no hay ningún Teatro Cartesiano.

Estos procesos de fijación de contenidos espacial y temporalmente distribuidos en el cerebro se pueden localizar con precisión en el tiempo y en el espacio, pero su inicio *no* marca el comienzo del contenido de la conciencia. Siempre queda abierta la cuestión de si un contenido en particular discriminado de este modo acabará por aparecer como un elemento de la experiencia consciente, y es una confusión, como veremos, preguntarse *cuándo se hace consciente.* Estas discriminaciones de contenido distribuidas producen, con el tiempo, algo *bastante parecido* a un flujo o secuencia narrativa, que puede considerarse sujeta a un proceso continuo de edición a través de muchos procesos distribuidos por el cerebro, que se prolonga indefinidamente hacia el futuro. Este flujo de contenidos se parece a un relato sólo a causa de su multiplicidad; en cualquier intervalo de tiempo existen múltiples «versiones» de fragmentos narrativos en diferentes estadios de edición y en diferentes puntos del cerebro.

Si sondeáramos este flujo en diferentes puntos del espacio o del tiempo se producirían efectos distintos, surgirían diferentes relatos por parte del sujeto. Si retrasamos demasiado este sondeo (toda la noche, por ejemplo), el resultado corre el riesgo de no ser ya una narración o, en su defecto, de ser una narración ya digerida o «racionalmente reconstruida» hasta el punto de carecer por completo de integridad. Si sondeamos «demasiado pronto» podemos obtener datos sobre cuán pronto el cerebro lleva a cabo una determinada discriminación, pero al precio de alterar lo que de otro modo sería la progresión normal del flujo múltiple. Más importante aún, el modelo de las Versiones Múltiples nos evita caer en la tentación de suponer que tiene que haber un único relato canónico (lo que podríamos llamar la versión «final» o «publicada»), es decir, el flujo *real* de la conciencia del sujeto, tanto si el investigador (o incluso el propio sujeto) puede acceder a él como si no puede.

Es probable que, por ahora, este modelo le parezca carente de sentido en tanto que modelo de la conciencia que usted conoce por su propia experiencia íntima. Ello se debe a que usted se siente todavía muy cómodo pensando que su conciencia se produce en el Teatro Cartesiano. Acabar con este hábito tan natural y cómodo, y convertir el modelo de las Versiones Múltiples en una alternativa creíble, exigirá un esfuerzo, un esfuerzo notable. Sin duda ésta será la parte más difícil del libro, pero es esencial para comprender la teoría en su conjunto y ¡no se puede prescindir de ella! Gracias a Dios, no se incluyen fórmulas matemáticas en estas páginas. Será suficiente con que usted piense con cuidado y con claridad, asegurándose de que se forma

una imagen correcta de las cosas y sin dejarse seducir por falsas imágenes. Hemos incluido un buen número de experimentos mentales que conducirán a su imaginación por este difícil camino. Prepárese pues para un ejercicio agotador. Al final del camino habrá usted descubierto una nueva visión de la conciencia, que comporta una reforma sustancial (pero no una revolución radical) de nuestra manera de pensar sobre el cerebro. (Para un modelo similar, véase el modelo de la conciencia de William Calvin (1989a, 1989b) basado en el concepto de «rotación de escenarios».)

Una buena manera de llegar a comprender una teoría nueva es verla en acción, dando cuenta de un fenómeno relativamente simple que se resistía a la vieja teoría. La muestra A es un descubrimiento sobre el movimiento aparente producido, me enorgullece poder decirlo, gracias a la pregunta de un filósofo. El cine y la televisión se basan en la creación de un movimiento aparente a partir de la presentación de una rápida sucesión de imágenes «fijas», y, desde los albores del cine, los psicólogos se han dedicado al estudio de este fenómeno que Max Wertheimer (1912), el primero en estudiarlo de forma sistemática, llamó *phi*. En el caso más simple, si dos pequeños focos, separados por no más de 4 grados de ángulo visual, se encienden por un breve espacio de tiempo en una rápida sucesión, parecerá como si un único punto luminoso se moviera hacia adelante y hacia atrás. Se han estudiado numerosas variaciones del fenómeno phi, la más sorprendente de las cuales es la referida por los psicólogos Paul Kolers y Michael von Grünau (1976). El filósofo Nelson Goodman había preguntado a Paul Kolers si el fenómeno phi persistía cuando los dos puntos de luz eran de colores diferentes y, si así era, qué ocurría en el color «del» punto de luz cuando «éste» se movía. ¿Desaparecería la ilusión del movimiento aparente, sustituida por la experiencia de dos focos encendiéndose por separado? ¿Cambiaría gradualmente de color un supuesto punto de luz en «movimiento», trazando una trayectoria a través del sólido de los colores (la esfera tridimensional que proyecta todos los tonos)? (Puede que usted quiera hacer sus propias predicciones antes de seguir leyendo.) La respuesta, cuando Kolers y von Grünau llevaron a cabo sus experimentos, fue inesperada: se encendieron dos focos de color durante 150 mseg cada uno (con un intervalo de 50 mseg); el primer foco parecía empezar a moverse para cambiar después de color de forma brusca *en la mitad de su ilusorio pasaje* hacia el segundo punto. Goodman se preguntó entonces: «¿Cómo somos capaces... de intercalar el punto de luz en el espacio-tiempo intermedio a lo largo del trayecto que va del primer destello al segundo destello *antes de que el segundo destello se haya producido*?» (Goodman, 1978, pág. 73).

Evidentemente, la misma pregunta es pertinente para cualquier variedad del fenómeno phi, pero el fenómeno phi de los colores descrito por Kolers es el que ilustra más claramente el problema. Supóngase que el primer foco es rojo y que el segundo es verde. A menos que haya «precognición» en el cerebro (una hipótesis bastante extravagante, cuya evaluación pospondremos *sine die*), el contenido ilusorio, *rojo-cambia-a-verde-a-medio-camino*, no

puede crearse hasta *después* de que el cerebro haya identificado el segundo punto de luz. Sin embargo, si el segundo punto de luz ya está «en la experiencia consciente», ¿no es ya demasiado tarde para intercalar el contenido ilusorio entre la experiencia consciente del punto de luz rojo y la experiencia consciente del punto de luz verde? ¿Cómo lleva a cabo el cerebro este juego de manos?

El principio de que las causas deben preceder a los efectos se aplica a los procesos múltiples distribuidos que llevan a cabo las tareas de edición en el cerebro. Todo proceso que necesita información de alguna fuente debe necesariamente esperar a recibir esa información; no puede estar ahí hasta que no llegue. De esta manera eliminamos toda posibilidad de dar una explicación «mágica» o precognitiva del fenómeno phi del cambio de colores. El contenido *punto de luz verde* no puede ser atribuido a ningún evento, consciente o inconsciente, hasta que la luz del foco verde haya alcanzado el ojo y desencadenado la actividad neuronal en el sistema visual hasta el nivel en que la discriminación del verde se lleva a cabo. Así pues, la discriminación (ilusoria) del rojo-convirtiéndose-en-verde debe producirse *después* de la discriminación del punto de luz verde. Sin embargo, dado que lo que usted experimenta conscientemente es *primero rojo, después rojo-convirtiéndose-en-verde, y finalmente verde*, se sigue que («seguramente») su conciencia del acontecimiento completo debe retrasarse hasta después de que el punto de luz verde haya sido percibido (¿inconscientemente?). Si usted encuentra atractiva esta conclusión, entonces sigue usted anclado en la idea del Teatro Cartesiano. Un experimento mental le ayudará a liberarse.

3. Revisiones orwellianas y estalinianas

> No estoy seguro de si los demás no me perciben o si, una fracción de segundo después de que mi rostro se cruzó en su horizonte, una millonésima de segundo después de que su mirada cayera sobre mí, ya han comenzado a borrarme de su memoria: olvidado antes de haber llegado a la escasez, triste arcángel de un recuerdo.
>
> ARIEL DORFMAN, *Máscaras*, 1988

Suponga que me pongo a manipular su cerebro y que consigo insertar en su memoria una ficticia mujer con sombrero allí donde no la había antes (por ejemplo, en la fiesta del domingo). Si, el lunes, cuando usted recuerda la fiesta, se acuerda de ella y no consigue encontrar ninguna fuente interna que le permita poner en duda lo que usted recuerda, seguiríamos estando en posición de decir que usted *nunca* la experimentó; en cualquier caso, no en la fiesta del domingo. Evidentemente, los recuerdos (ficticios) que usted experimentará después pueden ser todo lo claros que quiera: el martes todos estaremos de acuerdo en que usted ha tenido la clara experiencia consciente de que en la fiesta había una mujer con sombrero, pero la *primera*

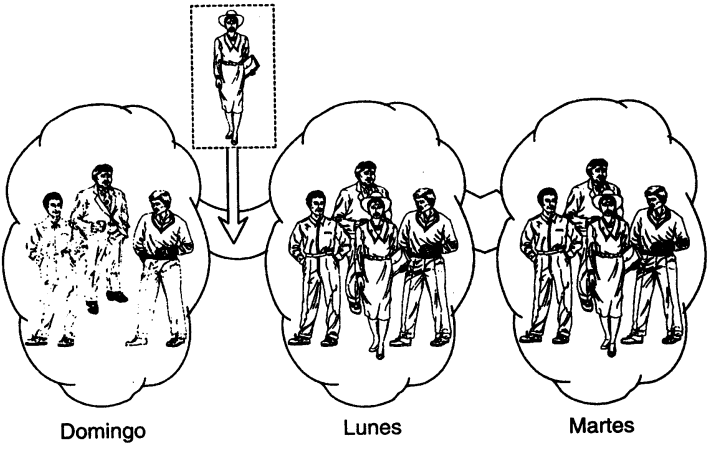


Figura 5.8

de estas experiencias, insisto, fue el lunes, y no el domingo (aunque a usted no le parezca que sea así).

No tenemos el poder de insertar recuerdos falsos mediante neurocirugía, pero a veces nuestros recuerdos no engañan, de manera que lo que no podemos conseguir por medios quirúrgicos ocurre en el cerebro por sí solo. A veces nos parece tener un recuerdo, incluso muy vívido, de experiencias que nunca han ocurrido. Llamaremos *orwellianas* a estas contaminaciones postexperienciales o revisiones de la memoria, en honor a la espeluznante visión de la novela *1984* de George Orwell, donde el Ministerio de la Verdad se ocupa hacendosamente de reescribir la historia, negando así el acceso al pasado (real) a las generaciones futuras.

La posibilidad de las revisiones postexperienciales (orwellianas) ilustra un aspecto de una de nuestras distinciones más fundamentales: la distinción entre lo que es apariencia y lo que es realidad. Habida cuenta de que reconocemos la posibilidad (al menos en principio) de que se produzcan revisiones orwellianas, reconocemos también el riesgo que corremos al inferir «esto es lo que realmente ocurrió» a partir de «esto es lo que yo recuerdo», y, por tanto, nos resistimos —con motivo— a cualquier «operacionalismo» diabólico que intente convencernos de que lo que recordamos (o lo que la historia refleja en sus archivos) *es precisamente* lo que ocurrió realmente.[4]

4. El operacionalismo es (aproximadamente) la idea o principio según el cual «si no se puede descubrir ninguna diferencia, entonces es que no hay diferencias», o, como a menudo se dice, «si grazna como un pato y anda como un pato, entonces es un pato». Para una reconsideración de los puntos débiles y los puntos fuertes del operacionalismo, véase Dennett (1985a).

La revisión orwelliana es una manera de tomarle el pelo a la posteridad. La otra manera es organizar falsos procesos, preparando cuidadosamente falsos testimonios y confesiones, y completándolo todo con falsas pruebas. A esta táctica la llamaremos *estaliniana*. Nótese que si normalmente estamos seguros del tipo de falsificación que se ha intentado sobre nosotros, sea orwelliana o estaliniana, ello no es más que un afortunado accidente. Ante cualquier campaña de desinformación *que tuviera éxito*, si nos preguntáramos si lo que aparece en los periódicos son informes de procesos que nunca se produjeron, o informes verídicos de farsas de procesos que sí se produjeron, no seríamos capaces de ver la diferencia. Si *todos* los indicios —periódicos, cintas de vídeo, recuerdos personales, inscripciones en las lápidas de un cementerio, testigos vivos— fueran eliminados o falseados, no tendríamos ninguna manera de saber si *primero* se produjo una manipulación y después se llevó a cabo un proceso ante un tribunal cuyas actas tenemos ahora ante nosotros, o si, por el contrario, *después* de una ejecución sumaria se construyó una historia relatando todos los hechos, es decir, no se llevó a cabo *en realidad* ningún proceso.

La distinción entre los métodos orwellianos y estalinianos de falsificación de archivos funciona perfectamente en la vida cotidiana, a escalas de tiempo macroscópicas. Por ello podríamos pensar que sigue aplicándose *en todos los casos*, pero eso es una ilusión, lo cual podemos demostrar acto seguido por medio de un experimento mental que se diferencia del que acabamos de considerar únicamente en la escala de tiempo.

Suponga que está usted parado de pie en una esquina cuando pasa corriendo delante suyo una mujer con el pelo largo. Alrededor de un segundo *después* de este episodio, el recuerdo oculto de otra mujer —una mujer con el pelo corto y gafas— contamina el recuerdo de lo que usted acaba de presenciar. Cuando un minuto más tarde se pregunta sobre algunos deta-

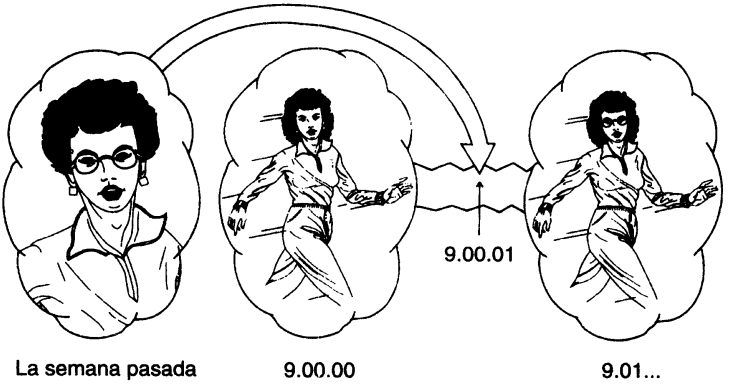


Figura 5.9

lles en relación a la mujer que usted vio, menciona, sincera pero erróneamente, sus gafas. Igual que en el caso de la mujer con sombrero de la fiesta, nos sentimos inclinados a decir que su experiencia *visual* original, al contrario que su recuerdo de la misma unos segundos después, *no* fue la de una mujer con gafas. Sin embargo, como resultado de la subsiguiente contaminación de la memoria, tiene usted la impresión de haberse fijado en sus gafas en el primer momento en que la vio. Se ha producido una revisión orwelliana: hubo un instante efímero, antes de que se produjera la contaminación de su memoria, en el que a usted *no* le parecía que llevara gafas. Durante ese momento, la *realidad* de su experiencia consciente era una mujer con el pelo largo y *sin* gafas, pero este hecho histórico ha quedado anulado; no ha dejado huella, gracias a la contaminación de la memoria que se produjo un segundo después de haber visto a la mujer.

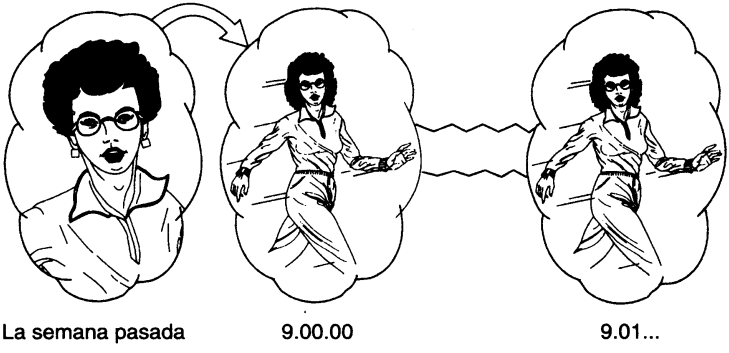


Figura 5.10

Esta interpretación de lo ocurrido se ve comprometida, sin embargo, por una explicación alternativa. Sus recuerdos ocultos de esa mujer con gafas podrían haber contaminado con igual facilidad su experiencia *en el camino de subida*, es decir, durante el procesamiento de la información que se efectúa «antes de la conciencia», de modo que usted en realidad tuvo una *alucinación* con las gafas desde el principio mismo de su experiencia. En este caso, el primer recuerdo obsesivo de una mujer que llevaba gafas le ha gastado una broma estaliniana al crear un proceso falso en la experiencia, que usted recuerda con precisión más tarde, gracias a un registro de su memoria. Para el sentido común ambos casos son todo lo diferentes que pueden ser: contado de la primera manera (figura 5.9), usted no sufre ninguna alucinación en el momento en que pasa la mujer, pero con posterioridad sufre alucinaciones de la memoria; usted tiene recuerdos falsos sobre su experiencia «real». Contado de la segunda manera (figura 5.10), usted alucina en el momento en que la mujer pasa, y después recuerda con claridad la alucina-

ción (que «realmente se produjo en la conciencia»). ¿Son éstas realmente dos posibilidades distintas, independientemente de cómo dividamos el tiempo?

No. En este caso, la distinción entre revisiones perceptivas y las revisiones de memoria que tan bien funciona a otras escalas ya no nos ofrece ninguna garantía de tener sentido. Hemos entrado en esa área brumosa donde el punto de vista del sujeto se distribuye por el espacio y el tiempo, y la pregunta *¿es orwelliana o estaliniana?* pierde su fuerza.

Hay una ventana temporal que se abrió cuando la mujer del pelo largo pasó corriendo, excitando sus retinas, y que se cerró cuando usted expresó —a usted mismo o a otra persona— la convicción de que llevaba gafas. En algún punto de ese intervalo, el contenido *que lleva gafas* se añadió al contenido *mujer con el pelo largo.* Podemos asumir (y quizá más tarde confirmar con detalle) que hubo un breve plazo de tiempo en el que el contenido *mujer con el pelo largo* ya había sido discriminado en el cerebro *antes* de que erróneamente se le «asociara» el contenido *que lleva gafas.* Evidentemente, resulta plausible suponer que fue esa discriminación de la mujer con pelo largo lo que desencadenó el recuerdo anterior de la mujer con gafas. Lo que no podremos saber, sin embargo, es si esa asociación espúrea se produjo «antes o después del hecho en cuestión» —el presunto hecho de «la experiencia consciente real». ¿Era usted primero consciente de una mujer con el pelo largo y sin gafas y después consciente de una mujer con el pelo largo y con gafas, experiencia esta última que borró de su memoria la experiencia anterior, o estaba la experiencia consciente matizada desde el primer instante con las gafas?

Si el materialismo cartesiano fuera cierto, esta pregunta debería tener una respuesta, incluso si no pudiéramos —ni nosotros ni usted— establecerla retrospectivamente por medio de ningún procedimiento de evaluación. Y eso es porque el contenido que «cruzó la línea de meta el primero» fue o bien *mujer con el pelo largo* o bien *mujer con el pelo largo y con gafas.* Pero todos los teóricos insistirán en afirmar que el materialismo cartesiano es falso. Lo que no han sabido reconocer, sin embargo, es que ello comporta la aceptación de que no hay ninguna línea de meta, y por tanto que el orden temporal de discriminaciones no puede ser lo que fija el orden subjetivo de la experiencia. Esta conclusión no es fácil de aceptar, pero podemos intentar resaltar su atractivo examinando las dificultades que uno encuentra cuando se empeña en aferrarse a la alternativa tradicional.

Considérese el fenómeno phi de los colores estudiados por Kolers. Los sujetos *atestiguan* haber visto el punto de luz en movimiento cambiar de color a medio camino de su trayectoria entre el rojo y el verde. Kolers tuvo la ingeniosa idea de potenciar la definición de esta pequeña porción de texto utilizando un puntero que los sujetos debían «superponer», retrospectivamente-pero-tan-pronto-como-fuera-posible, sobre la trayectoria del punto de luz en movimiento aparente; al utilizar el puntero, los sujetos ejecutaban un acto de habla cuyo contenido era: «el punto de luz cambió de color precisamente aquí» (Kolers y von Grünau, 1976, pág. 330).

Así pues, en el mundo heterofenomenológico de los sujetos, hay un cambio de color en el punto medio de la trayectoria, y la información sobre a qué color hay que cambiar (y en qué dirección hay que moverse) tiene que venir de alguna parte. Recuérdese cómo plantea Goodman el problema: «¿Cómo somos capaces... de intercalar el punto de luz en el espacio-tiempo intermedio a lo largo del trayecto que va del primer destello al segundo destello *antes de que el segundo destello se haya producido*?». Algunos teóricos han pensado que quizá la información proviene de una *experiencia previa*. Quizás, igual que los perros de Pávlov que esperaban que se les diera de comer cada vez que oían un timbre, los sujetos han llegado a esperar ver el segundo punto de luz cada vez que ven el primero, y, por la fuerza del hábito, se representan el paso antes de recibir la información sobre el caso particular. Esta hipótesis ha sido refutada, sin embargo. Incluso en la primera prueba (es decir, sin que exista la posibilidad de que se haya producido condicionamiento), las personas experimentan el fenómeno phi. Además, en intentos subsiguientes se puede cambiar el color del segundo foco o la dirección del movimiento al azar sin que por ello desaparezca el efecto. Así pues, la información sobre el segundo punto de luz (su color y localización) debe ser utilizada de alguna manera por el cerebro a fin de crear la versión «corregida» que refieren los sujetos.

Considérese, primero, la hipótesis de que nos hallamos ante un mecanismo estaliniano: en la sala de edición del cerebro, situada antes de la conciencia, se produce un retraso, un pequeño bucle como el que realiza la cinta de las emisiones «en directo» y que concede a los censores en la sala de realización unos segundos para tapar con pitos las obscenidades antes de emitir la señal al exterior. *En la sala de edición*, llega primero el fotograma A, el del punto de luz rojo, y entonces, cuando llega el fotograma B, el del punto de luz verde, se crean algunos fotogramas intermedios (C y D) que se montan intercalados en la película (en el orden A, C, D, B) antes de proyectarla en el teatro de la conciencia. Para cuando el «producto acabado» llega a la conciencia, ya lleva la inserción ilusoria.

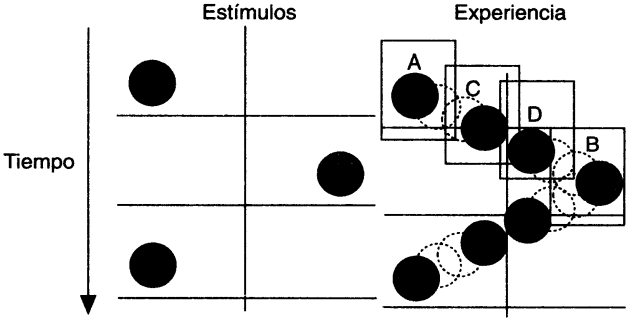


Figura 5.11

Como alternativa, tenemos la hipótesis de que hay un mecanismo orwelliano: poco después de la conciencia del primer punto de luz y del segundo punto de luz (sin ilusión del movimiento aparente alguna), una especie de historiador revisionista, en la estación de llegada de la biblioteca de la memoria en el cerebro, observa que la historia lisa y llana en este caso no tiene mucho sentido, así que interpreta los acontecimientos, rojo-seguido-de-verde, construyendo una narración sobre el paso intermedio, incluido el cambio de color, e instala la historia resultante de incluir los fotogramas C y D (en la figura 5.11) en la biblioteca de memoria para toda referencia futura. Como trabaja rápido, en una fracción de segundo —la cantidad de tiempo que se necesita para construir (pero no proferir) un testimonio oral de lo que se ha experimentado—, el registro sobre el que usted se basa, almacenado en la biblioteca de la memoria, ya está contaminado. Usted *dice* y *cree* que vio el movimiento aparente y el cambio de color, pero eso no es más que una alucinación de la memoria, y no un recuerdo preciso de su conciencia original.

¿Cómo podemos saber cuál de las dos hipótesis es la correcta? Parece que podemos rechazar la hipótesis estaliniana con bastante facilidad a causa del retraso en la conciencia que postula. En el experimento de Kolers y von Grünau, había una diferencia de 200 mseg entre el comienzo del punto rojo y el del punto verde, y dado que, *ex hypothesi*, la *experiencia completa* no puede componerse en la sala de edición hasta que el contenido *punto de luz verde* ha alcanzado la sala de edición, la conciencia del punto de luz rojo debería retrasarse como mínimo otro tanto. (Si la sala de edición enviara inmediatamente el contenido punto de luz rojo al teatro de la conciencia, antes de recibir el fotograma B y de fabricar los fotogramas C y D, el sujeto presumiblemente experimentaría una laguna en la película, un retraso de al menos 200 mseg entre A y C, lo cual es tan perceptible como un vacío de la longitud de una sílaba en una palabra, o como la ausencia de cinco fotogramas en una película.)

Supóngase que pedimos a los sujetos que opriman un botón «tan pronto como experimenten el punto de luz rojo». Apenas hallaríamos diferencia en cuanto al tiempo de respuesta ante sólo un punto rojo que ante un punto rojo seguido 200 mseg después por un punto verde (en cuyo caso los sujetos atestiguan la existencia de movimiento aparente y cambio de color). ¿Puede ello ser debido a que en la conciencia *siempre* hay un retraso de 200 mseg por lo menos? No. Existen abundantes datos que demuestran que las respuestas bajo control consciente, pese a ser más lentas que respuestas tales como el parpadeo reflejo, se producen cerca de las latencias (retrasos) mínimas que son físicamente posibles. Una vez restados los tiempos de recorrido de las cadenas de impulsos de entrada y de salida, así como el tiempo de preparación de la respuesta, no queda tiempo de «procesamiento central» suficiente con el que ocultar un retraso de 200 mseg. Por consiguiente, las respuestas por presión de un botón deberían haberse iniciado antes que la discriminación del segundo estímulo, el del punto de luz verde.

Esto parece conceder la victoria a la hipótesis orwelliana, un mecanismo de revisión postexperiencial: tan pronto como el sujeto es consciente del punto de luz rojo, inicia el acto de presionar el botón. *Mientras está llevando a cabo este acto,* se hace consciente del punto de luz verde. *Entonces,* ambas experiencias se borran de la memoria y se sustituyen por el registro revisionista del punto rojo moviéndose y volviéndose verde a medio camino. Acto seguido atestigua, sincera *pero falsamente,* haber visto el punto rojo moviéndose hacia el punto verde antes de cambiar de color. Si el sujeto insiste en afirmar que fue realmente consciente desde el principio de que el punto rojo se movía y cambiaba de color, el teórico orwelliano le explicará firmemente que está equivocado; su memoria lo está engañando; el hecho de que presionara el botón cuando lo hizo constituye una evidencia concluyente de que era consciente del punto de luz rojo (estacionario) antes de que el punto verde se encendiera. Después de todo, sus instrucciones eran presionar el botón *cuando fuera consciente* del punto rojo. Tiene que haber sido consciente del punto rojo unos 200 mseg antes de haber sido consciente de su movimiento y de su cambio de color. Si a él no le parece que es así, se debe a que está simplemente equivocado.

El defensor de la alternativa estaliniana no se sentirá derrotado por esto, sin embargo. En realidad, insiste, el sujeto respondió al punto rojo *antes* de ser consciente de él. Las instrucciones que se le dieron (responder al punto rojo) se han colado de alguna manera desde la conciencia a la sala de edición, la cual (inconscientemente) inició el acto de presionar el botón antes de enviar la versión corregida para ser «visionada» (los fotogramas ACDB) hacia la conciencia. La memoria no ha engañado al sujeto; nos está refiriendo exactamente aquello de lo que fue consciente, con la excepción de su insistencia en afirmar que presionó conscientemente el botón después de ver el punto rojo; su acto «prematuro» de presionar el botón se desencadenó de forma inconsciente (o preconsciente).

Allí donde la teoría estaliniana postula una reacción de presión del botón ante una detección inconsciente del punto rojo, la teoría orwelliana postula una experiencia consciente del punto rojo que es destruida inmediatamente por lo que la sigue en la memoria. Éste es, pues, el problema: tenemos dos modelos diferentes de lo que ocurre en el fenómeno phi de los colores. Uno propone una «suplantación» estaliniana en el camino de subida, el camino preexperiencial, mientras que el otro propone una «revisión de la memoria» orwelliana en el camino de bajada, el camino postexperiencial, y *ambos* son consistentes con *cualquier cosa* que diga, piense o recuerde el sujeto. Nótese que la incapacidad para distinguir entre las dos posibilidades no es aplicable únicamente a los *observadores exteriores,* a quienes se les puede suponer una falta de datos privados a los que el sujeto tiene un «acceso privilegiado». Usted, en calidad de sujeto en un experimento del fenómeno phi, *no podría* descubrir nada en la experiencia desde su propia perspectiva de la primera persona que favoreciera una teoría sobre la otra; la experiencia «le parecerá la misma» en cualquiera de los casos.

¿Es realmente así? ¿Si prestara realmente atención a su experiencia, no sería usted capaz de notar la diferencia? Suponga que el investigador se lo pusiera fácil, ralentizando el proceso y alargando gradualmente el intervalo de tiempo entre los estímulos del punto de luz rojo y el del verde. Es evidente que si el intervalo es lo bastante largo usted podrá ver la diferencia entre *percibir* el movimiento e *inferir* el movimiento. (Es una noche oscura y tormentosa; con el primer relámpago me ve usted a su izquierda; dos segundos más tarde hay otro relámpago y usted me ve a su derecha. Usted inferirá que yo me he movido; no hay ninguna duda de que en este caso usted está únicamente infiriendo el movimiento, y no *viendo* mi desplazamiento.) A medida que el investigador prolonga el intervalo entre estímulos, llegará un momento en que usted será capaz de efectuar esta discriminación. Ahora usted diría algo parecido a lo siguiente:

> «En esta ocasión no me pareció que el punto rojo se hubiera movido, pero después de ver el punto verde, pensé que el punto rojo se había movido y había cambiado de color».

De hecho, existe una gama intermedia de intervalos para los cuales la fenomenología es un tanto paradójica: ¡se perciben los puntos como dos focos estacionarios y como una sola cosa que se mueve! Este tipo de movimiento aparente es fácilmente distinguible del movimiento aparente más rápido y uniforme que vemos en las películas y en la televisión, pero nuestra capacidad de establecer *esta* discriminación no es relevante para resolver el conflicto entre los teóricos orwelliano y estaliniano. Ambos están de acuerdo en que usted puede establecer esta discriminación bajo las condiciones adecuadas. En lo que no están de acuerdo es en cómo describir los casos de movimiento aparente que usted *no puede* distinguir del movimiento real: aquellos casos en los que usted realmente *percibe* un movimiento ilusorio. Someramente, ¿en estos casos, su memoria lo está engañando, o son sus ojos los que lo están engañando?

Pero incluso si usted, el sujeto, no puede decidir si el fenómeno es estaliniano u orwelliano, ¿no podrían los científicos, en tanto que observadores exteriores, encontrar algo en su cerebro que resolviera el problema en una dirección u otra? Algunos quisieran rechazar esto como *inconcebible*. «¡Imagine a alguien que sepa mejor que usted mismo de qué es usted consciente! ¡Imposible!» Pero, ¿es esto realmente inconcebible? Analicémoslo con más detenimiento. Supongamos que los científicos dispusieran de información realmente precisa (conseguida gracias a diferentes tecnologías de escáner cerebral) sobre el «momento de llegada» exacto o la «creación» de toda representación, o vehículo de contenido, en cualquier parte de su sistema nervioso. Ello les permitiría establecer el momento *más temprano* en que podría usted reaccionar —consciente o inconscientemente— ante un contenido particular (excluyendo la precognición milagrosa). Sin embargo, el instante *real* en que usted *fuera consciente* de ese contenido (en caso de que lo fuera

en algún momento) se produciría un poco más tarde. Usted debería ser consciente de ello lo bastante pronto como para explicar la inclusión por su parte del contenido en cuestión en algún acto de habla recordatorio posterior —asumiendo que, por definición, cualquier entidad de su mundo heterofenomenológico es una entidad de su conciencia—. De tal modo se fijaría el momento *más tardío* en que el contenido «se hizo consciente». Pero, como hemos visto, si ello nos deja con un espacio de tiempo de unos cientos de milisegundos durante el cual la conciencia de la entidad en cuestión tuvo que producirse, y si hay muchas entidades diferentes que deben atravesar la ventana temporal (el punto rojo y el punto verde; la mujer de pelo largo con y sin gafas), no hay manera de utilizar sus *testimonios* para ordenar la representación de los acontecimientos en la conciencia.

Sus testimonios orales retrospectivos deben ser neutrales con respecto a dos presuntas posibilidades, pero ¿no podrían encontrar los científicos otros datos que pudieran utilizar? Podrían si existiera algún buen motivo para afirmar que cierto tipo de conducta no verbal (abierta o interna) es un claro indicador de conciencia. Aquí es, precisamente, donde el razonamiento falla. Ambos teóricos coinciden en aceptar que *no* hay ninguna reacción comportamental ante un contenido, con la excepción del acto posterior de referir, que no sea una mera reacción inconsciente. En el modelo estaliniano hay un acto inconsciente de presionar un botón (¿y por qué no?). Ambos teóricos también coinciden en aceptar que puede haber una experiencia consciente que no tenga ningún efecto comportamental. En el modelo orwelliano hay conciencia momentánea del punto de luz rojo estacionario, lo cual no deja rastros en ninguna otra reacción posterior (¿y por qué no?).

Ambos modelos pueden dar cuenta de *todos* los datos sin problemas; y no sólo los datos de que disponemos ahora, sino cualquier dato que podamos imaginarnos que pueda surgir en el futuro. Ambos dan cuenta de los testimonios orales: una teoría dice que son ingenuamente erróneos, mientras que la otra dice que son testimonios precisos de errores experimentados. Asimismo, podemos suponer que ambos teóricos tienen *exactamente* la misma teoría de lo que ocurre en su cerebro; están de acuerdo sobre dónde y cuándo el contenido erróneo entra en la vías causales del cerebro; simplemente no coinciden al determinar si tal localización debe considerarse pre o postexperiencial. Ambos proponen la misma explicación de los efectos no verbales, con una pequeña diferencia: uno dice que son el resultado de contenidos discriminados inconscientemente, mientras que el otro dice que son el resultado de contenidos discriminados conscientemente pero olvidados. Finalmente, ambos dan cuenta de los datos subjetivos —todo lo que puede obtenerse a partir de la perspectiva de la primera persona—, porque están incluso de acuerdo en cómo les debe «parecer» a los sujetos: éstos no deberían ser capaces de distinguir entre experiencias ilegítimas y experiencias mal memorizadas de inmediato.

Así pues, a pesar de las primeras apariencias, existe solamente una diferencia de orden verbal entre ambas teorías (para un diagnóstico similar, véase

Reingold y Merikle, 1990). Ambos teóricos nos cuentan exactamente la misma historia, con la única discrepancia del lugar donde sitúan la mítica gran divisoria, un punto en el tiempo (y, por tanto, también un punto en el espacio) cuya localización *precisa y exacta* es imposible con la ayuda de los sujetos, pero también cuya localización es neutral con respecto a todos los demás rasgos de sus teorías. Ésta es una diferencia que no permite establecer ninguna diferencia.

Considérese una analogía contemporánea. En el mundo editorial es normal distinguir entre lo que es la revisión de errores previa a la publicación, y lo que es la corrección de «erratas» con posterioridad a la publicación. Sin embargo, en el mundo académico las cosas se hacen mucho más rápido hoy en día gracias a la comunicación electrónica. Con el advenimiento de los procesadores de textos, los programas de autoedición y el correo electrónico, a menudo se da el caso de que circulan simultáneamente muchas versiones distintas de un mismo artículo, con el autor revisando y corrigiendo a medida que recibe comentarios y sugerencias por correo electrónico. El acto de fijar el momento de la publicación, y pasar así a considerar como texto *canónico* —el texto que acabará por registrarse, el que se citará en las bibliografías— una de las versiones del artículo, se convierte hasta cierto punto en un hecho arbitrario. Con frecuencia ocurre que la mayoría de los lectores a quien va dirigido el texto, es decir, aquellos cuya opinión cuenta, leen solamente una primera versión; la versión «publicada» queda como material de archivo. Si lo que estamos buscando es provocar efectos importantes, entonces la mayoría de los efectos importantes, si no todos, de escribir un artículo se reparten entre todas las numerosas versiones del mismo; antes solía ocurrir que la mayor parte de los efectos importantes de un artículo se producía *después* de su publicación en una revista y *a causa de* su publicación. Ahora que los distintos candidatos para la publicación no poseen importancia funcional alguna, debemos tomar una decisión arbitraria sobre lo que vale como publicación del texto, en caso de que realmente necesitemos tal distinción. No hay ninguna culminación o punto de inflexión en el camino que va del borrador al archivo.

Asimismo —y ésta es la consecuencia fundamental del modelo de las Versiones Múltiples—, si queremos fijar un instante del procesamiento en el cerebro como el instante de la conciencia, la decisión será arbitraria. Siempre se puede «trazar una línea» en el flujo de procesamiento en el cerebro, pero no hay ninguna diferencia funcional que permita motivar la calificación de ajustes inconscientes o preconscientes para todos aquellos estadios previos, y la calificación de contaminaciones postexperienciales de la memoria para todas las correcciones del contenido posteriores (tal y como lo revelan los recuerdos). La distinción se desvanece cuando se examina desde cerca.

4. Retorno al teatro de la conciencia

La regla de oro del astrónomo:
si no lo has escrito,
entonces es que no ha ocurrido.
CLIFFORD STOLL, *The Cuckoo's Egg*, 1989

Como cualquier libro sobre prestidigitación le enseñará, los mejores trucos se acaban antes de que los espectadores piensen que han empezado. Llegados a este punto, es posible que usted piense que acabo de jugarle una buena. He argumentado que a causa de la expansión espaciotemporal del punto de vista del observador en el cerebro, la evidencia no nos permite distinguir entre las teorías orwelliana y estaliniana de la experiencia consciente y que, por tanto, *no hay tal diferencia*. Esto es una especie de operacionalismo o verificacionismo que no considera la posibilidad de que existan hechos decisivos sobre la cuestión que sean inaccesibles a la ciencia, incluso cuando la ciencia incorpora la heterofenomenología. Además, la verdad es que parece obvio que estos hechos *sí que existen*, que nuestra experiencia consciente inmediata consiste precisamente en estos hechos.

Convengo en que, efectivamente, parece obvio; si no fuera así, no debería preocuparme tanto en este capítulo por demostrar que lo que parece tan obvio es falso a pesar de todo. De lo que parezco haber prescindido, con cierta premeditación, es de algo similar al tan ridiculizado Teatro Cartesiano de la conciencia. Puede que usted sospeche que bajo esta fachada de *antidualismo* («¡vamos a quitarnos de encima esta cosa fantasmal!»), yo haya hecho desaparecer (literalmente) algo sobre lo que Descartes tenía razón: existe una localización funcional de algún tipo donde las entidades de la fenomenología se... *proyectan*.

Ha llegado el momento de enfrentarse a esta sospecha. Nelson Goodman plantea el problema cuando, refiriéndose al experimento de Paul Kolers sobre el fenómeno phi de los colores, afirma que «no parece dejarnos otra elección que aceptar una teoría de la construcción retrospectiva o creer en la clarividencia» (Goodman, 1978, pág. 83). Debemos rechazar la clarividencia, así que ¿en qué consiste exactamente la «construcción retrospectiva»?

> Tanto si consideramos que la percepción del primer punto de luz se retrasa, se conserva o se recuerda, llamaré a esta concepción teoría de la construcción retrospectiva: la teoría según la cual la construcción de lo percibido como algo que se produce entre los dos puntos de luz no tiene lugar antes del segundo punto de luz.

En un primer momento, Goodman parece dudar entre una teoría estaliniana (la percepción del primer punto de luz se retrasa) y una teoría orwelliana (la percepción del primer punto de luz se conserva o se recuerda), pero lo más importante es que su presunto revisionista (sea orwelliano o estali-

niano) no se limita a ajustar los juicios, sino que *constituye* material para *rellenar* las lagunas:

> ...cada uno de los puntos intermedios en la trayectoria entre los dos puntos de luz se rellena... con uno de los dos colores que se encienden y no con colores intermedios sucesivos. (Pág. 85.)

Goodman no tiene en cuenta la posibilidad de que el cerebro no tenga realmente que tomarse la molestia de «rellenar» nada con ninguna «construcción», ya que no hay nadie que esté mirando. Como hace explícito el modelo de Versiones Múltiples, una vez se ha llevado a cabo una discriminación, ésta no tiene que volver a producirse; el cerebro se limita a adaptarse a la conclusión a que se llega, elaborando una nueva interpretación de la información disponible para modular la conducta subsiguiente.

Goodman considera la teoría, que él atribuye a Van der Waals y Roelofs (1930), de que «el movimiento intermedio se produce retrospectivamente, se construye sólo después de que se encienda el segundo foco y *se proyecta hacia atrás* en el tiempo [la cursiva es mía]» (págs. 73-74). Todo ello parece apuntar hacia una visión estaliniana, pero con un sesgo que no augura nada bueno: se realiza una película final que luego se proyecta en un proyector mágico cuyo haz de luz viaja hacia atrás en el tiempo hasta la pantalla de la mente. Tanto si es esto lo que Van der Waals y Roelofs tenían en mente cuando propusieron la «construcción retrospectiva» como si no, probablemente es lo que llevó a Kolers (1972, pág. 184) a rechazar su hipótesis, insistiendo en afirmar que toda construcción se efectúa en «tiempo real». ¿Por qué debería el cerebro preocuparse por «producir» el «movimiento intermedio»? ¿Por qué no puede simplemente *llegar a la conclusión de que hubo* movimiento intermedio, e insertar la conclusión retrospectiva en el flujo de procesamiento? ¿No es esto suficiente?

¡Alto! Aquí es donde debe estar produciéndose el juego de manos (si es que lo hay). Desde el punto de vista de la tercera persona, he postulado la existencia de un sujeto, el sujeto heterofenomenológico, una especie de «a quien le corresponda» ficticio a quien, desde fuera, le atribuimos correctamente la creencia de que el movimiento intermedio ha sido experimentado. Así es como le parecería a *ese* sujeto (que no es más que la ficción de un teórico). Pero, ¿acaso, no hay también un sujeto *real*, en cuyo beneficio el cerebro debe organizar un espectáculo, rellenando todas las lagunas? Esto es lo que Goodman parece suponer cuando habla del cerebro llevando a cabo el acto de rellenar todos los puntos intermedios de la trayectoria. ¿En beneficio de quién se hacen estos dibujos animados? Para los espectadores del Teatro Cartesiano. Pero *como no hay tal teatro, tampoco existen los espectadores.*

El modelo de Versiones Múltiples coincide con Goodman en suponer que el cerebro crea retrospectivamente el contenido (el juicio) de que hubo movimiento intermedio, y que ese contenido queda entonces para regir las acti-

vidades posteriores y para dejar su huella en la memoria. Pero el modelo de Versiones Múltiples va más lejos al afirmar que el cerebro no se toma la molestia de «construir» representaciones con el objeto de «rellenar» lagunas. Eso sería malgastar el tiempo y (¿podemos decirlo?) *pintura*. El juicio *ya esta ahí*, y el cerebro puede ocuparse de otras tareas.[5]

La frase de Goodman sobre una «proyección hacia atrás en el tiempo» es equívoca. Podría querer decir algo más modesto y defendible: concretamente, que en el contenido se incluye una *referencia a un instante pasado en el tiempo*. De acuerdo con esta interpretación sería una afirmación equivalente a «esta novela nos lleva a los tiempos de la antigua Roma...», que nadie interpretaría, de un modo metafísicamente extravagante, como la afirmación de que la novela es una especie de máquina del tiempo. Ésta es la única lectura que es consistente con las demás concepciones que defiende Goodman, pero Kolers aparentemente consideró que significaba algo metafísicamente radical: que se producía una proyección real de algo desde un tiempo hacia otro tiempo.

Como veremos en el próximo capítulo, la confusión provocada por esta lectura radical del término «proyección» ha influido negativamente en la interpretación de otros fenómenos. Esta misma curiosa metafísica afectó la manera de pensar sobre la representación del espacio. En los tiempos de Descartes, Thomas Hobbes parece haber pensado que la luz, después de alcanzar el ojo, provocaba algún tipo de movimiento en el cerebro, lo cual hacía que algo *rebotara* de vuelta hacia el mundo exterior.

> La causa del sentido es el cuerpo exterior, u objeto, que impresiona el adecuado órgano sensorial, ya inmediatamente, como ocurre con el gusto y el tacto, ya mediatamente, como sucede con la vista, el oído y el olfato. Este estímulo, a través de los nervios y de otras ligaduras y membranas del cuerpo, continúa hacia adentro hasta llegar al cerebro y al corazón. Y allí causa una resistencia o contrapresión, o empeño del corazón por liberarse a sí mismo, empeño que, al estar dirigido hacia *afuera*, parece que es una materia externa. (*Leviatán*, Parte I, cap. 1, «Del sentido».)*

Después de todo, ahí es donde vemos los colores —¡en la superficie de los objetos![6] En la misma línea de razonamiento, podríamos suponer que

5. Existe una región en el córtex denominada MT, que responde al movimiento (y al movimiento aparente). Supóngase entonces que una cierta actividad en MT es la conclusión del cerebro de que hubo movimiento intermedio. Desde el punto de vista del modelo de Versiones Múltiples no tiene sentido preguntarse si ésa es una conclusión pre o postexperiencial. En otras palabras, sería un error preguntarse si la actividad en MT fue una «reacción a la experiencia consciente» (por parte de un historiador orwelliano) o si, por el contrario, fue una «decisión de representar el movimiento» (por parte del editor estaliniano).

* Trad. cast. de Carlos Mellizo, Alianza, Madrid, 1989, págs. 19-20. [N. del T.]

6. De hecho, el propio Hobbes era consciente de los problemas que comporta esta idea: «Porque si esos colores y sonidos estuvieran en los cuerpos u objetos que los causan, no podrían separarse de ellos como de hecho vemos que pueden separarse mediante lentes, o, en el caso del eco, por reflexión. Y aunque a cierta distancia el mismo objeto real parece estar investido de

cuando se machaca el pulgar, ello produce unas señales que viajan hasta los «centros del dolor» en el cerebro que, a su vez, «proyectan» el dolor *de vuelta al pulgar que es donde tiene que estar.* Después de todo, es ahí donde sentimos el dolor.

Todavía en la década de los cincuenta, esta idea seguía tomándose en serio, hasta el punto de hacer que J. R. Smythies, un psicólogo británico, se sintiera obligado a escribir un artículo para refutarla definitivamente.[7] La proyección de que hablamos cuando nos referimos a estos fenómenos *no* comporta relanzar algún efecto hacia el espacio físico, y no creo que ya nadie crea que esto es así. Sin embargo, los neuropsicólogos y los psicólogos, así como los teóricos de la acústica que diseñan sistemas estereofónicos, hablan a menudo de este tipo de proyección, y es lícito que nos preguntemos qué quieren decir con ello, si no es algún tipo de transmisión física de un lugar (o tiempo) a otro. ¿Qué comporta esta proyección? Analicémoslo con detalle a partir de un caso simple:

> Gracias a la colocación de los altavoces y al balance del volumen de sus emisiones respectivas, el oyente *proyecta* el sonido resultante de la soprano hacia un punto situado a medio camino entre ambos altavoces.

¿Cómo debemos interpretar este párrafo? Examinémoslo con detenimiento. Si los altavoces están sonando a todo volumen en una habitación vacía, no hay proyección. Si un oyente está en la habitación (un observador con unos buenos oídos y un buen cerebro), se produce la «proyección», lo cual *no* significa que el oyente emita algo hacia ese punto a medio camino entre ambos altavoces. No hay ninguna propiedad física de ese punto o del área vecina que se vea alterada por la presencia del oyente. En pocas palabras, y esto es lo que queremos decir al afirmar que Smythies tenía razón, que no hay proyección en el espacio de ninguna propiedad visual o auditiva. ¿Qué es lo que ocurre entonces? Bueno, *al observador le parece* que el sonido de la soprano proviene de este punto. ¿Y qué comporta este *parecerle al observador*? Si nuestra respuesta es que «comporta la proyección por parte del observador del sonido hacia ese punto en el espacio», entonces volvemos a

---

la fantasía que él engendra en nosotros, el objeto siempre será una cosa, y la imaginación o fantasía será otra» (*Leviatán*, mismo capítulo [pág. 20 de la edición castellana; N. del T.]). Sin embargo, este pasaje es susceptible de diferentes interpretaciones.

7. Smythies (1954). Esta pieza heroica es una buena demostración de cuán difícil era pensar sobre estas cosas hace tan sólo treinta y siete años. Smythies refuta enérgicamente una versión del manual de la teoría de la proyección; en la conclusión cita aprobatoriamente un párrafo de Bertrand Russell en el que se rechaza la misma idea: «Aquel que acepte la teoría causal de la percepción se verá abocado a concluir que los perceptos se hallan en nuestras cabezas, ya que llegan al término de una cadena causal de eventos físicos conducentes, espacialmente, del objeto al cerebro del que percibe. No podemos suponer que, al final de este proceso, el último efecto salte repentinamente hacia atrás hasta el punto de partida como cuando soltamos una cuerda elástica en tensión» (Russell, 1927).

estar donde estábamos, evidentemente, así que la gente intenta introducir alguna novedad diciendo, por ejemplo, algo así como que «el observador proyecta el sonido en un *espacio fenoménico*». Parece que hemos hecho algún progreso. Hemos rechazado la proyección en el espacio físico y la hemos trasladado a un espacio fenoménico.

¿Y qué es el espacio fenoménico? ¿Un espacio físico en el cerebro? ¿El espacio escénico del teatro de la conciencia situado en el cerebro? Literalmente no. Pero, ¿y metafóricamente? En el capítulo anterior, en el ejemplo de las «imágenes mentales» que manipulaba Shakey, expusimos un método para dar sentido a estos espacios metafóricos. En un sentido estricto pero metafórico, Shakey dibujaba formas en el espacio, se fijaba en determinados puntos de ese espacio, y basaba sus conclusiones en lo que descubría sobre esos puntos en el espacio. Pero ese espacio no era más que un espacio *lógico*. Era como el espacio del Londres de Sherlock Holmes, el espacio de un mundo ficticio, pero de un mundo ficticio anclado sistemáticamente en acontecimientos físicos reales que se producían en el espacio ordinario del «cerebro» de Shakey. Si tomáramos las proferencias de Shakey como expresión de sus «creencias», entonces podríamos decir que era el espacio en que *creía* Shakey, lo cual no lo convierte en algo real, como tampoco el hecho de que alguien crea en Fenhomo convierte a Fenhomo en alguien real. Ambos no son más que objetos intencionales.[8]

Así pues, tenemos *una manera* de dar un sentido a la idea de espacio fenoménico en tanto que espacio lógico. Éste es un espacio en el cual o hacia el cual nada se proyecta literalmente; sus propiedades se constituyen gracias a las creencias del sujeto (heterofenomenológico). Cuando decimos que el oyente proyecta el sonido hacia un punto en este espacio, lo que queremos decir es *solamente* que a él le parece que es de ahí de donde viene el sonido. ¿No es esto suficiente? ¿O acaso estamos pasando por alto una doctrina «realista» del espacio fenoménico en el que esa apariencia real puede ser proyectada?

Hoy por hoy nos sentimos bastante seguros con la distinción entre localización espacial en el cerebro del vehículo de la experiencia, y la localización «en el espacio experiencial» de la entidad experimentada. En breve, distinguiremos entre representante y representado, entre vehículo y contenido.*

8. «Es como si nuestro fenhominista, convertido ahora en fenhominólogo, se aferrara, en su confusión, a la estratagema desesperada de inventar un espacio de los dioses, un cielo, para que su amado Fenhomo pueda residir en él, un espacio que sea lo bastante real como para satisfacer al que cree en él, pero lo bastante remoto y misterioso como para ocultar al Fenhomo de los escépticos. El espacio fenoménico es el cielo de las imágenes mentales, pero si las imágenes mentales resultan ser reales, podrán residir fácilmente en el espacio físico de nuestros cerebros, y si, por el contrario, resultan no ser reales, podrán residir, con Papá Noel, en el espacio lógico de la ficción.» Dennett (1978a), pág. 186.

* Como podrá observar el lector, en este capítulo y el siguiente se introduce una distinción fundamental en el modelo de Dennett: la distinción entre vehículo y contenido. El autor, intencionadamente, evita utilizar el término «representación» cuando se refiere al vehículo de un contenido y utiliza en cambio la palabra inglesa *representing*. A fin de mantener tal distinción he-

Hemos alcanzado un grado de sofisticación suficiente como para reconocer que los productos de la percepción visual no son, literalmente, imágenes en la cabeza aunque *lo que éstas representan* es algo que las imágenes representan muy bien: la disposición en el espacio de diversas propiedades visibles. Deberíamos establecer la misma distinción para el tiempo: *el momento* en que se produce una experiencia en el cerebro debe distinguirse del momento en que ésta parece producirse. Efectivamente, como ha sugerido el psicolingüista Ray Jackendoff, lo que debemos llegar a comprender aquí no es más que una extensión banal de nuestras ideas tradicionales sobre la experiencia del espacio. La representación del espacio en el cerebro no siempre hace uso de el-espacio-en-el-cerebro para representar el espacio, y la representación del tiempo en el cerebro no siempre hace uso de el-tiempo-en-el-cerebro. Tan infundada es la idea del proyector de diapositivas espaciales que Smythies no podía encontrar en el cerebro, como lo es la idea del proyector de cine temporal que sugiere la lectura radical que hace Goodman de la frase «proyección hacia atrás en el tiempo».

¿Por qué las personas sienten la necesidad de postular la existencia de este proyector de apariencias? ¿Por qué se sienten inclinadas a pensar que no es suficiente con que las salas de edición en el cerebro se limiten a insertar contenidos en el flujo que conduce a la modulación de la conducta y la memoria? Quizá porque quieren mantener la distinción entre realidad y apariencia para la conciencia. Quieren resistirse a ese operacionalismo diabólico que dice que cuanto ha ocurrido (en la conciencia) es simplemente aquello que usted recuerda que ha ocurrido. El modelo de las Versiones Múltiples convierte al acto de «escribir en la memoria» en el criterio mismo para la conciencia; en ello consiste precisamente que lo «dado» sea «tomado» —tomado de una manera más que de otra. No hay realidad de la experiencia consciente independientemente de los efectos que ejercen los diversos vehículos de contenido sobre los actos subsiguientes (y, por tanto, sobre la memoria). Esto se parece demasiado al temido operacionalismo, y quizás mimamos en secreto al Teatro Cartesiano de la conciencia como el lugar en que todo cuanto ocurre «en la conciencia» ocurre *realmente*, tanto si con posterioridad ello es recordado correctamente como si no. Supóngase que algo ocurriera en mi presencia, pero que dejara su huella en mí durante sólo «una millonésima de segundo», como en el epigrama de Ariel Dorfman. ¿Qué significaría decir que fui consciente de ello, por muy breve y poco efectivo que fuera este estado? Si en alguna parte hubiera un Teatro Cartesiano privilegiado, como mínimo significaría que *la película que allí se pasó era muy buena*, incluso si nadie recuerda haberla visto. (¡Ya estamos otra vez!)

---

mos optado por utilizar las palabras «representante» y «representado» para referirnos al vehículo y al contenido, respectivamente; de este modo capturamos el evidente paralelismo que existe entre la concepción defendida por Dennett y la concepción del signo lingüístico como la unión de un *significante* y un *significado*, que, en un plano ligeramente distinto, defendió el lingüista suizo F. de Saussure. Queremos agradecer al profesor Dennett el habernos ayudado a aclarar este punto. [N. del T.]

Puede que el Teatro Cartesiano sea una imagen reconfortante porque mantiene la distinción entre apariencia y realidad en el corazón de la subjetividad humana, pero además de no estar motivado científicamente, es metafísicamente dudoso, ya que crea la extravagante categoría de lo objetivamente subjetivo, la manera en que a usted, real y objetivamente, le parece que son las cosas, ¡incluso cuando a usted no le parece que le parezca que son así (Smullyan, 1981). Algunos pensadores mantienen una oposición tan cerril al «verificacionismo» y al «operacionalismo», que se empeñan en negar su utilidad incluso en un ámbito en el que sí tiene sentido: el reino de la subjetividad. Lo que Clifford Stoll denomina la regla de oro del astrónomo es un comentario jocoso sobre la vaguedad de la memoria y los estándares de la evidencia científica, pero se convierte en una gran verdad cuando se aplica a lo que «se escribe» en la memoria. Podríamos calificar, por tanto, al modelo de las Versiones Múltiples de operacionalismo como *de la primera persona*, ya que niega tajantemente y por principio la posibilidad de conciencia de un estímulo en ausencia de la creencia por parte del sujeto en dicha conciencia.[9]

Cuando uno se opone a este operacionalismo, es normal que apele a hechos posibles que son inaccesibles para el test del operacionalista, pero ahora el sujeto es el propio operacionalista, así que la objeción se vuelve contra él: «El mero hecho de que usted no pueda decir, de acuerdo con sus preferencias, si fue o no fue consciente de *x*, no significa que usted no lo fuera. ¡Quizá sí que fue usted consciente de *x*, pero no puede hallar ninguna evidencia de ello!». ¿Acaso hay alguien que, después de unos momentos de reflexión, realmente quiera decir algo así? Esos supuestos hechos sobre la conciencia que se alejan nadando, fuera del alcance de los observadores «exteriores» e «interiores», son hechos verdaderamente extraños.

La idea se resiste a morir. Considérese la naturalidad de una frase como la siguiente: «He juzgado que las cosas eran así, porque así es como a mí me pareció que eran». Aquí se nos invita a pensar en términos de dos tipos distintos de estados o eventos: el parecer-de-una-determinada-manera y el subsiguiente (y consecuente) juzgar-que-es-de-esa-manera. Pudiera pensarse que el problema con el modelo de Versiones Múltiples para el fenómeno phi de los colores, por ejemplo, es que, pese a incorporar el fenómeno del juicio del sujeto de que hubo movimiento intermedio, no contempla —de hecho, niega explícitamente— la existencia de un evento que pueda ser considerado como la apariencia-de-que-hubo-movimiento-intermedio y sobre el que «se basa» el juicio en cuestión. En alguna parte debe de producirse «una presentación de evidencias», aunque sea en un falso proceso estaliniano, para que tales evidencias actúen como causa o fundamento del juicio.

Algunas personas suponen que la fenomenología apoya esta intuición. Tie-

9. El filósofo Jay Rosenberg me ha hecho notar que Kant considera acertada esta concepción, al afirmar que, en la experiencia, el *für mich* (el «por mí») y el *an sich* (el «en sí») son una misma cosa.

nen la impresión de que realmente se observan a sí mismos juzgando que las cosas son así *como resultado* del hecho de que a ellos les parece que son así. Nadie ha observado nunca algo así «en su fenomenología» porque un hecho sobre la causalidad como éste no sería observable (como ya señaló Hume hace mucho tiempo).[10]

Pregúntesele a un sujeto en un experimento del fenómeno phi de los colores si juzga que el punto rojo se movió y cambió de color porque a él le pareció que así había sido, o si le pareció que el punto se movía porque él lo juzgó así. Supongamos que el sujeto nos da una respuesta tan «elaborada» como la siguiente:

> Sé que no había *realmente* un punto en movimiento en el mundo —después de todo, no es más que movimiento aparente—, pero también sé que *pareció* que el punto se movía, así que además de mi juicio de que el punto pareció moverse, existe el evento *de que trata* mi juicio: el hecho de que el punto pareció moverse. No hubo movimiento real, así que debe haber una apariencia real de que el punto se movió sobre la cual se basa mi juicio.

Puede que el Teatro Cartesiano sea tan popular porque es el lugar en que las apariencias pueden producirse además de los juicios. Sin embargo, este argumento tan elaborado que se acaba de presentar es falaz. El postular una «apariencia real» *además* del juzgar o el «considerar» expresado en el testimonio del sujeto es multiplicar entidades sin necesidad. El tipo de presentación interna en que las apariencias reales se producen es una maniobra metafísica inútil, es una manera de nadar y guardar la ropa, especialmente porque los que tienden a pensar así no dudan en insistir en que dicha presentación interna no se produce en un espacio misterioso y dualista, henchido de un éter de fantasmas cartesianos. Cuando se rechaza el dualismo cartesiano, se tiene que rechazar el espectáculo que tendría lugar en el Teatro Cartesiano, así como la audiencia, porque ni el espectáculo ni la audiencia están en el cerebro, y el cerebro es el único objeto real donde podríamos buscarlos.

10. El filósofo Ned Block me contó una vez su experiencia como sujeto en un test de «lateralización». Tenía que fijar en línea recta la vista en un punto y de vez en cuando se encendía una palabra (o una no palabra como GHRPE) a la izquierda o a la derecha de ese punto. Su tarea consistía en presionar un botón si el estímulo era una palabra. Sus tiempos de reacción eran significativamente más largos para las palabras que aparecían en el área izquierda (y que, por tanto, accedían primero al hemisferio izquierdo), lo cual no hacía más que confirmar la hipótesis de que él, como la mayoría de las personas, tenía una fuerte lateralización para el lenguaje en el hemisferio izquierdo. Tampoco fue una sorpresa para el propio Block; lo que le pareció interesante fue «la fenomenología: las palabras que se encendían a la izquierda aparecían un poco borrosas». Le pregunté si creía que las palabras eran más difíciles de identificar porque eran borrosas, o si parecían borrosas porque eran más difíciles de identificar. Block admitió no encontrar una manera de distinguir entre estas dos explicaciones causales «opuestas» de su juicio.

5. El modelo de Versiones Múltiples en acción

Pasemos a revisar el modelo de Versiones Múltiples, ampliándolo un poco, y considerando con algo más de detalle la situación en el cerebro que lo justifica. En aras de la simplicidad, me concentraré en lo que ocurre en el cerebro durante la experiencia visual. Más adelante podremos extender el análisis a otros fenómenos.

Los estímulos visuales evocan cadenas de acontecimientos en el córtex que, gradualmente, dan a lugar a discriminaciones de especificidad cada vez mayor. En diferentes tiempos y lugares, se toman «decisiones» o se emiten «juicios» de diversa índole; literalmente, se causa la entrada de partes del cerebro en estados capaces de discriminar diversos rasgos, por ejemplo, primero el inicio del estímulo, después su localización, después su forma, luego el color (por un camino distinto), aún más tarde el movimiento (aparente) y, con el tiempo, el reconocimiento del objeto. Estos estados discriminativos localizados transmiten efectos a otros lugares, contribuyendo a nuevas discriminaciones y así sucesivamente (Van Essen, 1979; Allman, Meizin y McGuinness, 1985; Livingstone y Hubel, 1987; Zeki y Shipp, 1988). La pregunta natural, aunque un tanto inocente, que podemos hacernos es, ¿en qué lugar se reúne todo? Y la respuesta es, en ninguna parte. Algunos de estos estados distribuidos portadores de contenido pronto se desvanecen, sin dejar huella. Otros sí dejan huella, en posteriores testimonios orales de la experiencia y la memoria, en la «disponibilidad semántica» y en otras variedades del conjunto perceptivo, en el estado emocional, las inclinaciones comportamentales, etc. Algunos de estos efectos —por ejemplo, la influencia sobre los testimonios orales— son, cuando menos, síntomas de conciencia. Pero no hay ningún lugar en el cerebro por el cual deban pasar estas cadenas causales a fin de depositar su contenido «en la conciencia».

Tan pronto como se haya producido tal discriminación, ésta está disponible para provocar algún comportamiento, por ejemplo, la presión de un botón (o una sonrisa, o un comentario), o para modular algún estado informacional interno. Por ejemplo, la discriminación de la imagen de un perro podría crear un «conjunto perceptivo» —lo que haría temporalmente más fácil el ver perros (o incluso otros animales) en otras imágenes— o podría activar un dominio semántico determinado, lo que haría temporalmente que usted interpretara la palabra «ladrar» como el sonido emitido por este animal y no como la manera de hablar desagradable de algunos humanos. Como ya hemos señalado, este proceso, que avanza por múltiples vías, se ejecuta en unos cientos de milisegundos, durante los cuales se producen diversos añadidos, incorporaciones, enmiendas y sobreescrituras de contenido, en diferentes órdenes. Con el tiempo, éstos dan lugar a algo *bastante parecido* a un flujo o secuencia narrativa, que podemos concebir como sujeto a revisiones continuas por parte de diversos procesos distribuidos por el cerebro, y que se prolonga indefinidamente en el futuro. Los contenidos surgen, son revisados, contribuyen a la interpretación de otros contenidos o a la modu-

lación de la conducta (verbal o de otro tipo) y, durante el proceso, dejan sus huellas en la memoria, las cuales, con el tiempo, acaban por deteriorarse, por incorporarse a contenidos posteriores e incluso por ser sobreescritas por estos contenidos, total o parcialmente. Esta madeja de contenidos se asemeja a una narración únicamente por su multiplicidad; en cualquier intervalo de tiempo hay múltiples versiones de fragmentos narrativos en varios estadios de revisión, en varios puntos del cerebro. Mientras que algunos de los contenidos de estas versiones harán su efímera contribución para desaparecer después sin tener más efecto —algunos incluso no harán ninguna contribución—, otros seguirán jugando diferentes papeles en la modulación posterior de estados internos o de la conducta y aun algunos pocos sobrevivirán hasta el punto de hacer patente su presencia a través de notas de prensa publicadas en forma de conducta verbal.

Sondear este flujo en diferentes intervalos produce efectos distintos, causando narraciones diferentes en cada caso. Y éstas *son* verdaderas narraciones: versiones únicas de una porción de «el flujo de la conciencia». Si el sondeo se demora demasiado, es posible que el resultado no dé lugar a ninguna narración. Si se sondea «demasiado pronto», será posible recopilar datos sobre en qué momento tiene lugar en ese flujo un determinado proceso de discriminación, pero a costa de interrumpir el progreso normal del mismo.

¿Podemos decir que hay un «momento ideal para llevar a cabo el sondeo»? Si aceptamos la idea, bastante plausible por otra parte, de que dichas narraciones sufren un proceso uniforme de degradación, tanto por los detalles que desaparecen como por los adornos que se incorporan (lo que debería haber dicho en la fiesta tiende a convertirse en lo que dije en la fiesta), podemos justificar un sondeo relativamente temprano, muy poco después del estímulo que nos interesa en aquel momento. Sin embargo, también queremos evitar el interferir en el fenómeno con un sondeo prematuro. Habida cuenta de que la percepción se convierte imperceptiblemente en memoria, y la interpretación «inmediata« se convierte imperceptiblemente en reconstrucción racional, no podemos decir que haya un punto culminante válido para todo contexto en el cual podamos fijar nuestra atención para efectuar nuestros sondeos.

Aquello de lo que somos conscientes durante un intervalo de tiempo determinado no se define independientemente de los sondeos que utilizamos para causar una narración sobre ese período. Dado que las narraciones están sometidas a una revisión continua, no podemos decir que haya una narración única que valga como versión canónica, la «primera edición», en la cual se han vertido, definitivamente, los acontecimientos que han tenido lugar en el flujo de la conciencia del sujeto, y tal que cualquier desviación de esa versión deberá ser considerada como una corrupción del texto. Sin embargo, cualquier narración (o fragmento de narración) que ha sido causada fija una «línea del tiempo», una secuencia subjetiva de acontecimientos desde el punto de vista de un observador, que puede compararse con otras líneas del tiempo y, en particular, con la secuencia objetiva de acontecimientos pro-

ducida en el cerebro de ese observador. Como hemos visto, estas dos líneas *pueden* no superponerse en un registro ortogonal (bien alineadas): aunque la (mala) discriminación del *rojo-pasando-a-verde* se produjera en el cerebro *después* de la discriminación del *punto verde*, la secuencia *subjetiva* o *narrativa* es, evidentemente, *punto rojo, seguido de rojo-pasando-a-verde*, y *finalmente punto verde*. Así pues, dentro de la distribución temporal del punto de vista del sujeto pueden producirse diferencias de orden que den lugar a bucles temporales.

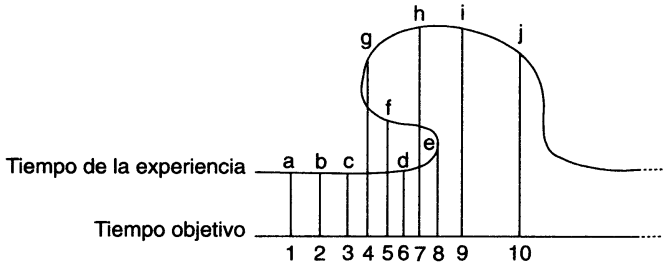


Figura 5.12

No hay nada de metafísicamente extravagante o problemático en este fallo de registro.[11] No es más misterioso o contrario a la causalidad que el caer en la cuenta de que las escenas de una película no siempre se ruedan en el mismo orden en que luego se montan, o que cuando uno lee la frase «Bill llegó a la fiesta después de Sally, pero Jane llegó antes que ambos», sabe de la llegada de Bill antes de saber de la llegada de Jane. El espacio y el tiempo del representante es un marco de referencia; el espacio y el tiempo de lo que el representante representa es otro. Este inocuo hecho metafísico sirve, no obstante, como fundamento de una categoría metafísica básica: cuando una porción del mundo pasa, de acuerdo con este método, a componer un ovillo de narraciones, dicha porción del mundo es un observador. En esto consiste precisamente el hecho de que haya un observador en el mundo, un algo que se siente ser.

Esto no es más que un esbozo de mi modelo alternativo. Todavía queda por aclarar en qué medida difiere del modelo del Teatro Cartesiano, demostrando cómo puede dar cuenta de determinados fenómenos. En el próximo capítulo pondremos el modelo a prueba con ciertas cuestiones bastante complejas, pero primero consideraremos algunos ejemplos más corrientes y simples que han sido objeto de discusión por parte de los filósofos.

Es posible que usted haya experimentado el fenómeno de conducir mu-

11. Esta manera de pensar me vino a la mente después de leer a Snyder (1988), aunque su manera de enfocar los problemas difiere ligeramente de la mía.

chos kilómetros enfrascado en una conversación con su acompañante (o en un soliloquio silencioso), para darse cuenta al final de que usted no recuerda la carretera, ni el tráfico, ni su actividad como conductor. Es como si hubiera estado conduciendo otra persona. Muchos teóricos (yo incluido, tengo que admitirlo —Dennett, 1969, págs. 116 y sigs.) han visto en ello un caso claro de «percepción inconsciente y acción inteligente». Pero, ¿era usted *realmente* inconsciente de todos los coches que pasaron, de los semáforos en rojo, de las curvas, durante todo el tiempo? Usted prestaba atención a otras cosas, pero es seguro que *si se le hubiera sondeado* sobre lo que *acababa* de ver, en diferentes momentos del viaje, usted habría tenido algún detalle, por somero que fuera, que referir. Es preferible interpretar el fenómeno de la «conducción inconsciente» como un caso de conciencia constante con repentinas pérdidas de memoria.

¿Es usted constantemente consciente del tic-tac del reloj? Si se parara de repente, lo notaría y podría decir sin dudarlo que se había parado; «usted no era consciente» del tic-tac hasta el momento en que se paró, y «nunca hubiera sido consciente de ello» si no se hubiera parado, pero ahora ocupa claramente un lugar en su conciencia. Un fenómeno aún más sorprendente es aquel en que usted es capaz de contar, retrospectivamente en la experiencia de la memoria, las campanadas de un reloj que ya había tocado cuatro o cinco campanadas cuando usted se dio cuenta. ¿Cómo es posible que usted *recuerde haber oído* con tanta claridad algo de lo que no era consciente desde el principio? La pregunta traiciona un compromiso con el modelo cartesiano; no hay hechos fijos sobre el flujo de la conciencia, independientemente de los sondeos determinados.

CAPÍTULO 6

TIEMPO Y EXPERIENCIA

> Puedo decir, claro está, que mis representaciones son sucesivas, pero esto sólo quiere decir que tenemos conciencia de ellas como situadas en una secuencia temporal, es decir, somos conscientes de ellas de acuerdo con la forma de nuestro sentido interno.
>
> IMMANUEL KANT, *Crítica de la razón pura*, 1781*

En el capítulo anterior consideramos un esbozo del modelo de las Versiones Múltiples que resuelve el problema de la «referencia hacia atrás en el tiempo», pero ignoramos algunas complicaciones mayores. En este capítulo seguiremos tratando estos asuntos, adentrándonos en territorios más peligrosos, examinando y resolviendo algunas controversias que han surgido entre los psicólogos y los investigadores del cerebro en relación a cómo se pueden explicar ciertos experimentos notoriamente problemáticos. Creo que es posible comprender el resto del libro sin necesidad de seguir con detalle los argumentos que se desarrollan en este capítulo, así que el lector puede limitarse a hojearlo o saltárselo si lo prefiere; sin embargo, he procurado organizar mi exposición tan claramente como he podido a fin de que los asuntos tratados aquí resulten comprensibles al profano, y además se me ocurren seis buenas razones para enfrentarse, después de todo, a las partes más técnicas.

1. Quedan por aclarar todavía muchas cosas de mi modelo de las Versiones Múltiples, y sólo viendo el modelo en acción podrá usted tener una visión más clara de su estructura.
2. Si todavía le quedan dudas sobre las diferencias que existen entre el modelo de las Versiones Múltiples, en tanto que teoría empírica, y el modelo tradicional del Teatro Cartesiano, estas dudas se disiparán ante el espectáculo que ofrecerán ambos modelos enfrentados.
3. Si usted se pregunta si no hago otra cosa que atacar a un hombre de paja, le resultará aleccionador el descubrir a ciertos expertos haciéndose un tremendo lío porque, muy a pesar suyo, son unos materialistas cartesianos.

* *Crítica de la razón pura* I, Primera parte: Estética trascendental, Sección 2.ª El tiempo, §7, A37-B54, Nota k. Traducción al castellano de Pedro Ribas, Alfaguara (Clásicos Alfaguara), Madrid, 1978. [N. del T.].

4. Si usted sospecha que he basado mi modelo en un único fenómeno, cuidadosamente escogido, como el fenómeno phi de los colores descrito por Kolers, podrá usted ver hasta qué punto se benefician muchos otros fenómenos de un tratamiento en términos de las Versiones Múltiples.
5. Muchos de los conocidos experimentos que examinaremos aquí han sido esgrimidos por *algunos* insignes expertos como refutaciones del tipo de teoría materialista que estoy presentando aquí, así que si debe haber un desafío *científico* a mi explicación de la conciencia, éste es el campo de batalla que la oposición ha escogido.
6. Finalmente, los fenómenos que trataremos son fascinantes y merecedores del esfuerzo que puede suponer aprender sobre ellos.[1]

I. Momentos efímeros y conejos saltarines

El testimonio oral que sigue a una experiencia suele ser una condición suficiente, aunque no necesaria, para considerar que efectivamente dicha experiencia se produjo. Es alrededor de este caso que giran todos los fenómenos más problemáticos. Suponga que, aunque su cerebro ha registrado —respondido a— ciertos aspectos de un evento, algo interviene entre esta respuesta interna y una ocasión subsiguiente que aprovecharía usted para relatar su testimonio. Si no hubiera tiempo u oportunidad para que ningún tipo de respuesta abierta (verbal o de otro tipo) incorporase alguna referencia a algún aspecto del primer evento, entonces se nos plantea el problema siguiente: ¿estos eventos no fueron percibidos conscientemente, o quizá fueron olvidados repentinamente?

Muchos experimentos se han ocupado de medir este «lapso de aprehensión». En un test de memoria acústica, usted tiene que escuchar una cinta en la que se ha grabado una serie de elementos sin relación, que se presentan con cierta rapidez (unos cuatro elementos por segundo, pongamos por caso), y que usted tiene que identificar. Usted no puede responder hasta que ha pasado toda la cinta y, cuando llega el momento de la identificación, identifica algunos de los elementos que oyó, pero no otros. Sin embargo, subjetivamente, usted escuchó cada uno de los elementos con la misma claridad. La pregunta que este hecho plantea es evidente: ¿de qué fue usted consciente entonces? No cabe duda de que toda la información de la cinta fue procesada por su aparato auditivo, pero ¿llegaron a la conciencia los rasgos identificativos de aquellos elementos que usted no nombró, o fueron simplemente registrados inconscientemente? *Parecen* haber estado ahí, en la conciencia, pero ¿estuvieron *realmente*?

En otro paradigma experimental, se le muestra una diapositiva en la que

1. Los argumentos y análisis presentados en este capítulo (y parte, también, de los desarrollados en el anterior) provienen de los trabajos descritos en Dennett y Kinsbourne (1992).

están impresas diversas letras. (Esto se hace con un *taquitoscopio*, un dispositivo de presentación de imágenes que puede ajustarse a fin de que el estímulo aparezca proyectado con un brillo determinado o durante un número determinado de milisegundos —a veces sólo 5 mseg, otras 500 mseg o más). Después, usted sólo es capaz de listar algunas de las letras, aunque no cabe duda de que también vio el resto. Usted insiste en que estaban allí, sabe exactamente cuántas había, y tiene la impresión de que se aprecian con claridad y nitidez. Y sin embargo no puede identificarlas. ¿Acaso las ha olvidado rápidamente, o quizá nunca llegó a percibirlas conscientemente desde el principio?

Un fenómeno tan bien estudiado como el del *metacontraste* (Fehrer y Raab, 1962), pone de manifiesto con claridad la idea principal que subyace al modelo de las Versiones Múltiples. (Para un examen de fenómenos similares, véase Breitmeyer, 1984). Si un estímulo se enciende en una pantalla, durante, pongamos por caso, unos 30 mseg (más o menos lo mismo que un fotograma de televisión) y va seguido inmediatamente por un estímulo que lo «enmascara», los sujetos *refieren* haber visto únicamente el segundo estímulo. El primer estímulo podría ser un disco coloreado y el segundo un anillo también coloreado en cuyo centro encajaría el disco que se presentó con anterioridad.

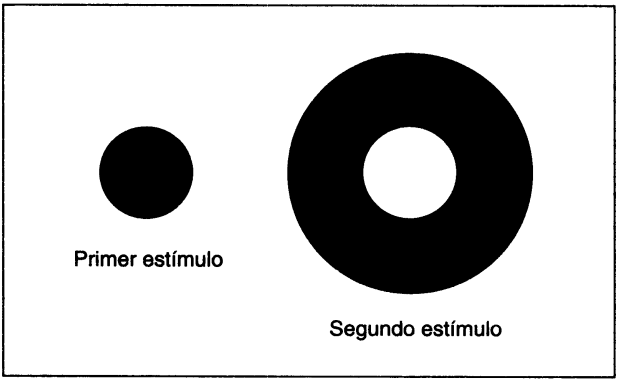


Figura 6.1

Si usted pudiera ponerse en el lugar del sujeto, podría comprobarlo por sí mismo; estaría usted dispuesto a jurar que sólo había un estímulo: el anillo. En la bibliografía psicológica, la descripción estándar de este fenómeno se realiza en términos estalinianos: el segundo estímulo, de algún modo *impide la experiencia consciente* del primer estímulo. En otras palabras, el segundo estímulo sale al paso del primer estímulo en su camino hacia la conciencia. Las personas muestran, no obstante, un alto grado de aciertos, que

no puede ser debido al azar, cuando se les hace adivinar si había uno o dos estímulos. Esto sólo es otra demostración, nos dice el teórico estaliniano, de que los estímulos pueden producir efectos en nosotros sin que seamos conscientes de ello. El primer estímulo nunca llega a salir al escenario de la conciencia, pero produce todos los efectos que tiene que producir de manera totalmente inconsciente. Podemos comparar esta explicación del metacontraste con su alternativa orwelliana: los sujetos sí que son conscientes del primer estímulo (lo cual explica su capacidad de adivinar correctamente que los estímulos eran dos), pero su recuerdo de la experiencia consciente ha sido eliminado *casi* por completo por el segundo estímulo (motivo por el cual niegan haberlo visto, a pesar de sus reveladoras intuiciones que van más allá del puro azar). El resultado es un punto muerto —y un motivo de azoramiento para ambas posturas—, ya que ninguna de las dos partes puede identificar un solo resultado experimental que *pudiera* darle la razón y resolviera así el conflicto.

Así es como el modelo de las Versiones Múltiples da cuenta del fenómeno del metacontraste. Cuando en un breve espacio de tiempo pasan muchas cosas, el cerebro tiene que hacer hipótesis simplificadoras. El contorno exterior de un disco se convierte rápidamente en el contorno interior de un anillo. El cerebro, informado inicialmente sólo de que algo ha ocurrido (algo con un perímetro circular en un lugar determinado), repentinamente recibe la confirmación de que, efectivamente, había un anillo, con sus contornos externo e interno. Sin más datos que apoyen la evidencia de que había un disco, el cerebro se inclina por la opción más conservadora de que sólo había un anillo. ¿Debemos insistir en el hecho de que sí que se experimentó el disco porque, *si el anillo no hubiera intervenido*, se habría señalado la presencia del primero? Ello comportaría cometer el error de suponer que podemos «congelar la imagen» en la película proyectada en el Teatro Cartesiano, y asegurarnos así de que el fotograma del disco consiguió llegar hasta el Teatro antes de que su recuerdo fuera eliminado por acontecimientos posteriores. El modelo de las Versiones Múltiples conviene en aceptar que la información sobre el disco ocupó, durante un breve espacio de tiempo, una posición funcional para contribuir así a un testimonio posterior, pero este estado pasó; no existe ninguna razón para suponer que este estado permaneció dentro del círculo encantado de la conciencia hasta que fue sobreescrito con nueva información, o, por el contrario, tampoco existen razones para pensar que nunca consiguió alcanzar este estado de privilegio. Las versiones compuestas en tiempos y lugares determinados del cerebro se retiran posteriormente de la circulación y son sustituidas por versiones revisadas, pero ninguna de ellas puede ser identificada como el contenido definitivo de la *conciencia*.

Una muestra aún más sorprendente de esta capacidad de revisión es el fenómeno del *conejo cutáneo*. Los psicólogos Frank Geldard y Carl Sherrick refirieron sus primeros experimentos en 1972 (véase también Geldard, 1977; Geldard y Sherrick, 1983, 1986). El brazo del sujeto descansa acolchado so-

bre una mesa, con unos martinetes colocados en dos o tres puntos a lo largo del brazo separados por unos treinta centímetros. Entonces se hace que los martinetes produzcan unos martilleos rítmicos, por ejemplo cinco en la muñeca, seguidos por dos cerca del codo y después tres más cerca del hombro. Estos martilleos se ejecutan con un intervalo entre estímulos de 50 mseg a 200 mseg. La cadena de martilleos puede durar, pues, menos de un segundo o, como máximo, dos o tres segundos. El sorprendente efecto que estos martilleos producen en los sujetos es que parecen desplazarse brazo arriba en una secuencia regular sobre puntos equidistantes, como si un pequeño animal estuviera subiendo por el brazo. En un primer momento, uno tiene la tentación de preguntarse *cómo sabía el cerebro* que después de los cinco martilleos en la muñeca, iba a haber más martilleos cerca del codo. Los sujetos ya experimentan la «partida» de los martilleos desde la muñeca con el segundo martilleo y, sin embargo, en las pruebas de preparación, en que los martilleos del codo y posteriores nunca se producen, los sujetos perciben los cinco martilleos en la muñeca de manera esperada. Es evidente que el cerebro no puede «saber» de un martilleo en el codo hasta que éste se haya producido. Si usted sigue hechizado por el Teatro Cartesiano, es posible que se sienta inclinado a conjeturar que el cerebro retrasa la experiencia consciente hasta el momento en que los martilleos han sido «recibidos» en un apeadero entre el brazo y el asentamiento de la conciencia (sea lo que sea) y, en este apeadero, se revisan los datos de acuerdo con alguna teoría del movimiento y se envían revisados a la conciencia. Pero, ¿retrasará siempre el cerebro la respuesta a un martilleo por si luego llegan más? Y si no lo hace, ¿cómo «sabe» cuándo hay que producir un retraso?

El modelo de las Versiones Múltiples demuestra que ésta es una pregunta descabellada. El desplazamiento en el espacio (a lo largo del brazo) es discriminado por el cerebro durante el tiempo en que se produce. También se discrimina el número de martilleos. Aunque en la realidad física los martilleos se concentran en puntos determinados, el supuesto más simple es que se distribuyen regularmente a lo largo de la duración espacio/temporal de la experiencia. El cerebro se relaja ante esta interpretación parca pero errónea *una vez* que se han registrado los martilleos, lo cual tiene el efecto de borrar cualquier interpretación (parcial) previa de los martilleos, aunque los efectos secundarios de estas interpretaciones pueden perdurar. Por ejemplo, supóngase que pedimos a los sujetos que presionen un botón cada vez que sientan *dos martilleos en la muñeca;* no nos sorprendería que pudieran iniciar el acto de presionar el botón *antes* de la discriminación de los martilleos del antebrazo que les inducirían a interpretar erróneamente que el segundo martilleo está desplazándose brazo arriba.

Debemos ser particularmente cuidadosos y no cometer el error de suponer que el contenido que derivaríamos a partir de un sondeo tan temprano constituye el «primer capítulo» del contenido de la narración con que nos encontraríamos si efectuáramos un sondeo del mismo fenómeno un tiempo después. Ello comporta la confusión de dos «espacios» distintos: el espacio

del representa*nte* y el espacio de lo representa*do*. Éste es un error tan tentador y tan común que merece que le dediquemos un apartado entero.

2. Cómo representa el tiempo el cerebro

El materialismo cartesiano, esa visión que nadie defiende pero en términos de la cual casi todo el mundo piensa, nos sugiere la siguiente imagen implícita. Sabemos que la información se desplaza por el cerebro y es procesada por diversos mecanismos en regiones distintas. Nuestras intuiciones sugieren que nuestro flujo de la conciencia consiste en eventos secuenciales, y que en cada instante cada elemento de esta secuencia puede clasificarse como si ya hubiese tenido lugar «en la conciencia» o como si no hubiese tenido lugar «aún». Y si es así, entonces (así parece) los vehículos que contienen contenidos y que se desplazan por el cerebro deben ser como vagonetas circulando por un carril; el orden en que pasan por algún punto será el orden en que «llegan» al teatro de la conciencia y (por tanto) «se hacen conscientes». Para determinar *el punto* en el cerebro en que se produce la conciencia, trácense todas las trayectorias de los vehículos de la información y véase por qué punto están pasando esos vehículos en el instante en que se hacen conscientes.

Un poco de reflexión sobre la tarea fundamental del cerebro nos hará ver dónde está el error en esta imagen. La tarea del cerebro es guiar al cuerpo que controla en un mundo de condiciones cambiantes y sorpresas repentinas, así que debe acumular información sobre este mundo y utilizarla *con rapidez* para «crear futuro», para producir anticipaciones a fin de andar un paso por delante del desastre (Dennett, 1984a, 1991b). Así pues, el cerebro tiene que representarse las propiedades temporales de los acontecimientos del mundo, y lo tiene que hacer con eficiencia. Los procesos que son responsables de la ejecución de esta tarea se distribuyen espacialmente por todo un cerebro sin un punto central, y la comunicación entre las regiones de este cerebro es relativamente lenta; los impulsos nerviosos electroquímicos viajan varios miles de veces más lentos que la luz (o que las señales electrónicas por un cable). El cerebro, por tanto, está sometido a una constante presión temporal. Con mucha frecuencia se ve obligado a arreglárselas para modular su salida en función de la entrada dentro de una ventana temporal que no permite descuidos que den lugar a retrasos. En el caso de los estímulos de entrada, hay tareas de análisis perceptivo, como la percepción del habla, que estarían por encima de los límites físicos de la maquinaria cerebral si ésta no utilizara ingeniosas estrategias anticipadoras que añaden redundancias a la entrada. El habla ordinaria se produce a una velocidad del orden de las cuatro o cinco sílabas por segundo, pero tan poderosas son las máquinas de análisis que hemos desarrollado para «segmentarla», que las personas son capaces de comprender «habla comprimida» —en la que las palabras se aceleran electrónicamente sin que por ello el tono se eleve hasta

sonar como el chillido de una ardilla— a velocidades que pueden superar las treinta sílabas por segundo. En el caso de la salida, muchas acciones pueden llevarse a cabo muy rápidamente, y se desencadenan con tanta precisión que el cerebro no tiene tiempo de ajustar sus señales de control en función de los estímulos de realimentación; actos tales como tocar el piano o tirar una piedra con puntería (Calvin, 1983, 1986) deben iniciarse *balísticamente.* (Los actos balísticos son como lanzar un misil *no* guiado; una vez se han desencadenado, sus trayectorias no pueden ser corregidas.)

¿Cómo puede entonces el cerebro seguir la pista de la información temporal que necesita? Considérese el siguiente problema: dado que la distancia entre un dedo del pie y el cerebro es mayor que la distancia de la cadera al cerebro, o del hombre al cerebro, o de la frente al cerebro, los estímulos lanzados simultáneamente en cada uno de estos lugares llegarán al cuartel general en envites sucesivos, si la velocidad a la que viajan es constante en todas las vías. ¿«Cómo» —se preguntará usted— «consigue el cerebro la simultaneidad central de representación necesaria para estímulos distales simultáneos»? Si se ocupara usted en una tarea de ingeniería especulativa inversa, podría pensar que quizá todos los canales nerviosos aferentes son como medidas de banda magnética enrollada en espiral como un muelle, todas de la misma longitud: los nervios de los dedos del pie están estirados al máximo, los de la frente están básicamente enroscados al cerebro. Las señales que viajan por el segundo canal entran en un largo bucle de esa espiral interna, saliendo de él para llegar al cuartel general sólo en el mismo instante en que lo hacen las señales que, sin haber sufrido ningún retraso, provienen del dedo del pie. O quizás usted pudiera imaginar que los canales nerviosos se estrechan a medida que se los estira (como rollitos de arcilla o como fideos hechos en casa), y que la velocidad del viaje varía a medida que el diámetro del canal disminuye. (Y así es en cierto modo, ¡pero en la dirección contraria, desgraciadamente! Las fibras nerviosas más gruesas conducen la información más rápido.) Éstos son modelos vívidos (y un poco tontos) de mecanismos que podrían resolver el problema, pero el error de principio es suponer que el cerebro necesita resolver *este* problema. El cerebro no tiene por qué resolver este problema, y por un obvio motivo de ingeniería: malgasta un tiempo precioso utilizando en toda la gama de operaciones que lleva a cabo un sistema de preferencias sensible a los «casos peores». ¿Por qué unas señales vitales procedentes de la frente (por ejemplo) deberían quedar retenidas en la antesala de la conciencia sólo porque en algún momento podría darse la situación de que ciertas señales concurrentes provenientes de los dedos de los pies tuvieran que converger con ellas?[2]

Los ordenadores digitales también dependen de estos retrasos para dar

2. Lo cual no significa que el cerebro nunca utilice «*buffers* de memoria» para amortiguar la interficie entre los procesos internos del cerebro y el asincrónico mundo exterior. La «memoria ecoica» donde se mantienen durante un tiempo ciertos patrones de estímulo mientras el cerebro empieza a procesarlos sería un buen ejemplo de ello (Sperling, 1960; Neisser, 1967; véase también Newell, Rosenbloom y Laird, 1989, pág. 107).

entrada a los casos peores y asegurar así la sincronía. El mecanismo de un circuito paralelo de adición que retiene sumas completadas hasta que un impulso de modulación temporal las libera sería un pariente próximo de esos imaginarios nervios en espiral. Los constructores de superordenadores deben ser extraordinariamente cuidadosos y asegurarse de que los cables que conectan las diversas partes de un circuito tengan la misma longitud, lo cual a menudo les obliga a incluir bucles de cable extra. Sin embargo, los ordenadores digitales se pueden permitir el lujo de esta ineficiencia localizada porque son terriblemente rápidos. (De hecho, la carrera que ha iniciado el mercado para obtener ordenadores cada vez más rápidos ha tenido como consecuencia que los ingenieros reconsideren estas minúsculas ineficiencias temporales; el motivo por el cual algunas de ellas siguen vigentes es que los ingenieros no saben cómo diseñar sistemas totalmente asíncronos, sin regulación mediante los impulsos de un reloj central.)

La imposición de una sincronía centralizada de las operaciones requiere la existencia de algunos retrasos. Actuando como ingenieros que trabajan a la inversa, podríamos pensar que si existieran mecanismos efectivos para que el cerebro represente la información sobre el tiempo que necesita a fin de evitar dichos retrasos, la evolución ya habría «dado con ellos». De hecho estos mecanismos existen, como podemos demostrar con un incidente histórico que, a una escala mayor, ilustra perfectamente el fenómeno, tanto en el espacio como en el tiempo.

Considérense las dificultades de comunicación con que se enfrentaba el gigantesco imperio británico antes de la invención de la radio y el telégrafo. Controlar un imperio mundial desde los cuarteles generales en Londres no era siempre factible. El incidente más célebre es sin duda el de la batalla de Nueva Orleans, el 8 de enero de 1815, quince días después de la tregua, firmada en Bélgica, que ponía fin a la guerra de 1812. Más de mil soldados británicos murieron en esa inútil batalla. Podemos utilizar la historia de esta debacle para ver cómo funcionaba el sistema. Supongamos que el día 1 se firmaba el tratado en Bélgica y que la noticia fue enviada por mar y tierra a América, India, África, etc. El día 15 tiene lugar la batalla en Nueva Orleans y las noticias de la derrota viajan por tierra y mar a Inglaterra, India, etc. El día 20, demasiado tarde, la noticia de la firma del tratado (y la orden de rendición) llega a Nueva Orleans. El día 35, supongamos, la noticia de la derrota llega a Calcuta, pero la noticia de la firma del tratado no llega hasta el día 40 (viajaba por vía terrestre, más lenta). Para el comandante en jefe británico de la India, la batalla «parecería» haberse librado antes de la firma del tratado, si no fuera por la práctica habitual de fechar las cartas, lo que le permitió efectuar las correcciones pertinentes.[3]

3. Me apresuro a añadir que estoy introduciendo algunos adornos históricos. Francis Rawdon-Hastings, el primer marqués de Hastings y segundo conde de Moira, era el gobernador de Bengala y comandante en jefe de la India en 1815, pero no tengo ni la más remota idea de cómo y cuándo fue informado del resultado de la batalla de Nueva Orleans.

Estos agentes separados por grandes distancias resolvieron la mayoría de sus problemas de comunicar información sobre el tiempo incorporando representaciones de la información temporal relevante en el contenido de sus señales, de modo que el momento de llegada a destino de las señales era *estrictamente irrelevante* para la información que éstas transportaban. Una fecha escrita en el encabezamiento de una carta (o registrada en el matasellos del sobre), proporciona al destinatario información sobre el momento en que ésta fue enviada, una información que sobrevive a cualquier retraso que se produzca en la llegada.[4] Esta distinción entre el tiempo representado (por el matasellos) y el tiempo del representante (el día de la llegada de la carta) es un ejemplo de la conocida distinción entre contenido y vehículo. Aunque esta solución en particular no está a la disposición de los agentes comunicadores del cerebro (porque no «conocen la fecha» en el momento de enviar el mensaje), el principio general de la distinción contenido/vehículo tiene una relevancia para los modelos de procesamiento de la información en el cerebro que nunca ha sido apreciada en su justa medida.[5]

En general, debemos distinguir los rasgos de los representa*ntes* de los rasgos de los representa*dos*. Uno puede chillar «¡Con cuidado, caminad de puntillas!» a pleno pulmón, hay imágenes enormes de objetos microscópicos, y no hay nada imposible en una pintura al óleo de un artista haciendo un boceto al carboncillo. La primera frase de una descripción escrita de un hombre de pie no tiene por qué describir su cabeza, ni la última frase tiene por qué describir sus pies. Este principio también se aplica, con menos claridad, al tiempo. Considérese la siguiente frase proferida oralmente: «Un brillante y breve destello de luz roja». Empieza con «un brillante» y acaba con «luz roja». Las porciones de dicha proferencia no son en sí mismas representaciones del principio o el final del breve destello rojo (Efron, 1967, pág. 714, hace una observación parecida). Ningún evento en el sistema nervioso puede tener una duración cero (ni tampoco puede tener una dimensión es-

4. Dicho «matasellos» podría, en principio, añadirse al vehículo de contenido en cualquier instante durante el viaje; si todo lo que llega a un punto determinado viene del mismo sitio, a través de la misma ruta y a la misma velocidad, el «tiempo de salida» del destino original puede marcarse retroactivamente, mediante la simple operación de restar una constante al tiempo de llegada al apeadero. Desde el punto de vista de la ingeniería, ésta es una posibilidad que el cerebro probablemente utiliza para efectuar ciertos ajustes automáticos en tiempos de viaje estándar.

5. Como señala Uttal (1979), dicha distinción *es* tenida en cuenta por los investigadores del cerebro: «La esencia de la mayor parte de las investigaciones que se han llevado a cabo en el campo de la codificación sensorial puede resumirse en una única idea particularmente importante: cualquier código posible es capaz de representar cualquier dimensión perceptiva; no es necesario que exista una relación isomórfica entre los datos neurológicos y psicofísicos. El espacio puede representar el tiempo, el tiempo puede representar el espacio, la localización puede representar la cualidad y, sin duda, las funciones neurológicas no lineales pueden representar con la misma precisión las funciones psicofísicas lineales y no lineales» (pág. 286). Sin embargo, pese a que la idea es conocida, enseguida veremos que algunos teóricos la comprenden comprendiéndola *mal*; la manera de «otorgarle un sentido» pasa por la reintroducción tácita de ese «isomorfismo» innecesario a través de un confuso proceso imaginario de traducción o «proyección» que se supone que tiene lugar en la conciencia.

pacial cero), así que todos tienen un principio y un final separados por un espacio de tiempo determinado. Si el evento mismo *representa* un evento en la experiencia, entonces el evento que representa debe tener también una duración distinta de cero, un principio, un final y una parte intermedia. No hay motivos para suponer, no obstante, que el principio del representante representa el principio del representado.[6] Aunque sí es cierto que diferentes mecanismos neurológicos extraen atributos diferentes en momentos diferentes (por ejemplo, localización frente a forma frente a color), y aunque si se nos pidiera que respondiéramos a la presencia de cada uno de ellos aisladamente lo haríamos con latencias diferentes, nosotros percibimos eventos, y no un torrente de elementos perceptivos o atributos analizados sucesivamente.[7]

Una novela o una narración histórica no tienen necesariamente que ser compuestas en el orden que eventualmente relatan; a veces los autores empiezan por el final y proceden hacia atrás. Además, una narración puede contener escenas retrospectivas o *flashbacks*, en los que los acontecimientos *se representan como* si hubieran ocurrido en un orden determinado mediante *representantes* que tienen lugar en órdenes distintos. De igual modo, la representación por parte del cerebro de A *antes que* B no tiene necesariamente que producirse de la siguiente manera:

Primero:

un representante de A,

Seguido por:

un representante de B.

La frase «B después de A» es un ejemplo de vehículo (oral) que representa el hecho de que A está antes que B, y el cerebro puede valerse de la misma libertad de localización temporal. Lo que importa al cerebro no es necesariamente cuándo se producen actos individuales de representación en diver-

6. Véase Pylyshyn (1979, pág. 278): «Nadie ... está dispuesto a hablar *literalmente* de propiedades físicas de eventos mentales tales como el color, el tamaño, la masa, etc... aunque *sí* hablamos de ellas como *representaciones* (o como poseedoras del contenido experiencial) de dichas propiedades. Por ejemplo, hablando con propiedad, nadie podría decir que un pensamiento (o una imagen) es grande o rojo, sino que es un pensamiento *sobre* algo grande o rojo (o que es la imagen de algo grande o rojo). (...) Por tanto, debería resultar sorprendente que hablemos con tanta libertad de la *duración* de un evento mental».

7. Como señala el psicólogo Robert Efron: «Cuando observamos por primera vez un objeto con visión central, no experimentamos súbitamente el objeto como nos aparecería con la visión más periférica, y después como nos aparecería con una visión menos periférica. (...) Similarmente, cuando cambiamos nuestra atención de un objeto de conciencia a otro, no hay experiencia de especificidad "creciente" del nuevo objeto de conciencia; simplemente percibimos un nuevo objeto» (1967, pág. 721).

sas partes del mismo (¡mientras éstos se produzcan a tiempo para controlar aquello que debe controlarse!), sino su *contenido temporal*. Es decir, lo que importa es que el cerebro pueda seguir controlando los acontecimientos «en el supuesto de que A ocurrió antes que B», independientemente de si la información de que A se ha producido entre en el sistema relevante del cerebro y sea reconocida como tal antes o después de la información de que B ha ocurrido. (Recordemos al comandante en jefe de Calcuta: primero fue informado de la batalla, y después fue informado de la tregua, pero dado que pudo extraer la información de que la tregua se produjo antes, pudo asimismo actuar en consecuencia. Debió, por tanto, *juzgar* que la tregua se produjo antes de la batalla; no tenía por qué organizar una especie de espectáculo de «reconstrucción histórica» en el que recibiese las cartas en orden «adecuado».)

Algunos han argumentado, sin embargo, que el tiempo es una de esas cosas que la mente o el cerebro *tiene que* representar «consigo mismo». El filósofo Hugh Mellor, en su libro *Real Time* (1981, pág. 8), plantea la cuestión con claridad y vehemencia:

> Supóngase por ejemplo que veo un evento *e* preceder a otro evento *e**. Primero tengo que presenciar *e* y después *e**, de tal modo que mi visión de *e* sea recordada en mi visión de *e**. Es decir, mi visión de *e* afecta a mi visión de *e**: esto es lo que hace que yo —correcta o erróneamente— vea *e* precediendo a *e** y no al revés. Pero ver *e* precediendo a *e** comporta ver primero *e*. Así pues, el orden causal de mi percepción de estos acontecimientos, al fijar el orden temporal que yo percibo que tienen, fija el orden temporal de las percepciones mismas. (...) Debemos hacer hincapié... en el hecho sorprendente de que la percepción del orden temporal necesita de percepciones temporalmente ordenadas. *Ninguna otra propiedad o relación necesita incorporarse así a las percepciones que uno tiene de ellas* (la cursiva es mía): la percepción de la forma y del color, por ejemplo no necesita tener a su vez forma o color.

Esto es falso, pero también tiene algo de cierto. Dado que la función fundamental de la representación en el cerebro es controlar la conducta en tiempo real, la organización temporal de los representantes es, *hasta cierto punto*, esencial para su tarea, y lo es por dos motivos.

Primeramente, al principio de un proceso perceptivo la organización temporal puede ser *lo que determine el contenido*. Considérese cómo distinguimos un punto que se mueve de derecha a izquierda de un punto que se mueve de izquierda a derecha en una pantalla de cine. La *única* diferencia entre ambos casos puede ser el orden temporal en que se proyectan dos fotogramas (o más). Si primero se proyecta A antes que B, se verá el punto moviéndose en una dirección; si primero se proyecta B antes que A, se verá el punto moviéndose en la dirección opuesta. La única diferencia en los estímulos que podría utilizar el cerebro para llevar a cabo la discriminación de direcciones es el orden en que se produjeron. Dicha discriminación es, por tanto, pura cuestión de lógica, basada en la capacidad del cerebro de llevar a cabo

una discriminación del orden temporal de una ciudad determinada. Puesto que los fotogramas de cine se presentan normalmente a una velocidad del orden de veinticuatro imágenes por segundo, sabemos que el sistema visual puede resolver el orden entre estímulos que se producen dentro de un margen de tiempo aproximado de 50 mseg. Ello significa que las propiedades temporales reales de las señales —el tiempo de inicio, su velocidad en el sistema y, por tanto, su tiempo de llegada— deben ser cuidadosamente controladas hasta que la discriminación haya tenido lugar. En caso contrario, la información sobre la que se debe basar la discriminación se perdería o se vería oscurecida

A una escala mayor, este fenómeno se produce en algunas ocasiones al principio de una regata; usted *ve* un barco cruzando la línea de salida y después *oye* el pistoletazo de salida: ¿llegó el barco a la línea demasiado pronto? La respuesta constituye una imposibilidad lógica, a menos que usted calcule la diferencia de tiempos de transmisión de la luz y el sonido hasta el punto en que usted llevó a cabo la discriminación. Una vez se ha emitido un juicio (*o todo va bien o el barco número 7 ha cruzado la línea demasiado pronto*), su contenido puede ser transmitido a los participantes de manera pausada, sin que haya necesidad de preocuparse por lo rápido o lo lejos que *tiene* que viajar para cumplir con su cometido.

Así pues, la organización temporal de algunas representaciones es importante *hasta el momento* en que se produce una discriminación como *de-izquierda-a-derecha* (o *cruzar la línea demasiado pronto*), pero una vez se ha producido, con la intervención de algún circuito en el córtex (o de algún observador en el barco de los árbitros), el contenido del juicio puede enviarse, con cierto descuido temporal, a cualquier punto en el cerebro donde esta información pueda ser utilizada. Sólo así podemos explicar lo que de otra manera es el enigmático hecho de que las personas en ocasiones son incapaces de emitir juicios temporales de muy poca precisión, mientras que en otras ocasiones son muy precisas con sus juicios sobre cuestiones que desde el punto de vista lógico exigen una mayor acuidad temporal (como ocurre con los juicios sobre la dirección del movimiento). Utilizan discriminadores especializados (y especialmente localizados) para producir juicios de gran calidad.

Ya mencionamos, aunque de forma somera, cuál es la segunda restricción sobre la organización temporal: no importa en qué orden se produzcan los representantes, con la condición de que se produzcan a tiempo para controlar las conductas adecuadas. La función de un representante puede depender del cumplimiento de un *plazo*, que es una propiedad temporal del vehículo que efectúa la representa*ción*. Este hecho es particularmente claro en entornos sometidos a una gran presión temporal como la Iniciativa de Defensa Estratégica concebida en EE.UU. recientemente. El problema no radica en saber cómo hacer que un ordenador represente, con gran precisión, el lanzamiento de un misil, sino en saber cómo representar el lanzamiento de un misil con gran precisión durante el breve espacio de tiempo en que

todavía se puede hacer algo para remediarlo. El mensaje de que un misil fue lanzado a las 6.04.23.678 AM EST puede representar con precisión y para siempre el momento del lanzamiento, pero su utilidad puede haberse desvanecido a las 6.05 AM EST.* Para toda tarea de control debe haber, pues, una *ventana de control temporal* dentro de la cual los parámetros temporales de los representantes puedan, en principio, ser alterados *ad libitum.*

Los plazos que ponen límite a estas ventanas no son fijos, sino que dependen de la tarea. Si, en vez de interceptar misiles, usted está escribiendo sus memorias o respondiendo a preguntas sobre el caso Watergate (Neisser, 1981), la información que usted necesite recuperar sobre la secuencia de acontecimientos de su vida a fin de controlar sus acciones puede ser recuperada prácticamente en cualquier orden, y además usted puede tomarse su tiempo para llevar a cabo las inferencias necesarias. O, por tomar un caso intermedio más próximo a los fenómenos que estamos considerando, suponga que usted va en un barco a la deriva y que quiere saber si se acerca o se aleja de ese peligroso arrecife que ve a lo lejos. Suponga que *ahora* usted conoce la distancia que lo separa del arrecife (porque, pongamos por caso, ha medido el ángulo que subtiende en su campo visual); a fin de responder a su pregunta, puede usted esperar un poco y volver a medir el ángulo, *o,* si media hora más tarde fotografiara el arrecife con su cámara Polaroid, usted podría medir el ángulo en la fotografía, efectuar algunos cálculos y, *retrospectivamente,* saber a qué distancia se encontraba en el momento de sacar la fotografía. Para poder emitir un juicio sobre la dirección en que usted se está moviendo, tiene que calcular dos distancias: la distancia a mediodía y la distancia a las 12.30, por ejemplo, pero a efectos prácticos no importa qué distancia calcule primero. Lo importante es que usted sea capaz de efectuar los cálculos con la suficiente rapidez para poder empezar a remar antes de que sea demasiado tarde.

Así pues, la represent*ación* del tiempo por parte del cerebro está anclada al tiempo mismo de dos maneras distintas: la organización temporal del representante puede ser lo que nos proporcione la evidencia o lo que determine el contenido, y toda la importancia de representar el tiempo de las cosas puede perderse si el representante no se produce a tiempo como para establecer la distinción que se quiere establecer. Espero que Mellor aprecie estos dos factores, y que los tuviera en mente en el momento de hacer las afirmaciones que cité más arriba, pero comete el error natural de pensar que la combinación de ambos restringe *completamente* la representación del tiempo, de modo que el orden de los representantes *siempre* refleja el orden del contenido. De acuerdo con este análisis, no hay lugar para la «distribución» temporal, mientras que aquí hemos defendido la necesidad de que *haya* distribución temporal —a pequeña escala— *porque* debe haber distribución espacial (a pequeña escala) del punto de vista del observador.

Las causas deben preceder a los efectos. Este principio fundamental ase-

* EST: Eastern Standard Time, la hora oficial en los Estados del Este de los EE.UU. [N. del T.]

gura que las ventanas de control temporal estén acotadas por ambos extremos: por el instante más temprano en que la información puede llegar al sistema, y por el instante más tardío en que la información puede tener alguna contribución causal al control de alguna conducta determinada. Todavía no hemos visto *cómo* podría utilizar el cerebro el tiempo disponible en una ventana de control para ordenar la información que recibe y convertirla en una «narración» coherente que se utiliza para regir las respuestas del cuerpo.

Entonces, ¿cómo pueden los procesos cerebrales inferir propiedades temporales? Los sistemas de «fechado postal» o de «matasellos» no son teóricamente imposibles, pero existe un mecanismo menos costoso, quizá menos infalible pero más plausible desde el punto de vista biológico: lo que podríamos denominar *dispositivo dependiente del contenido.* Una analogía útil sería la del estudio de sonido donde la banda sonora se «sincroniza» con la película. Los diversos segmentos de la cinta de audio pueden haber perdido todos sus marcadores temporales, de modo que no hay una manera simple y mecánica de disponerlos de manera que se correspondan con las imágenes. Sin embargo, al pasarlos en uno y otro sentido junto con la película, buscando puntos de coincidencia, se pueden hallar los ajustes «óptimos». La claqueta al principio de cada toma —«escena tres, toma siete, ¡acción!, CLAC»— proporciona un punto de referencia doble, visual y auditivo, cuya sincronización hace que el resto de la cinta y los fotogramas entren automáticamente en sincronía. Sin embargo, existen normalmente tantos puntos de referencia mutua que este punto de referencia convencional al principio de cada toma no es más que una cómoda redundancia. Que la grabación final sea la adecuada depende del *contenido* de la película y de la cinta de audio, pero no de un sofisticado análisis del contenido. Un editor que no hablara japonés encontraría difícil y aburrido, pero no imposible, el sincronizar una película en esta lengua. Además, el orden temporal de los estadios del proceso de ajustar las piezas en la grabación es independiente del contenido del producto; el editor puede organizar la escena tres antes de organizar la escena dos y, en principio, incluso podría hacer todo el trabajo pasando los segmentos «a la inversa».

Ciertos procesos «estúpidos» en el cerebro pueden llevar a cabo procesos similares de manipulación y encaje. Por ejemplo, el cómputo de la profundidad en un estereograma de puntos aleatorios (figura 5.7, página 126) es un problema espacial para el que fácilmente podemos imaginar análogos temporales. En principio, pues, el cerebro puede resolver sus problemas de inferencia temporal mediante tales procesos, sin obtener información de los ojos derecho e izquierdo, sino a partir de cualquier fuente de información que participe en un proceso que requiere juicios temporales.

De ello se siguen dos puntos importantes. Primero, dichas inferencias temporales pueden realizarse (o, dichas discriminaciones temporales pueden efectuarse) comparando el *contenido* (de bajo nivel) de las distintas matrices de datos, y este proceso en tiempo real no tiene por qué producirse en el orden

temporal representado por el resultado de dicho proceso. Segundo, una vez que dicha inferencia temporal se ha llevado a cabo, lo cual puede haber ocurrido *antes* de que otros procesos hayan extraído rasgos de alto nivel, no tiene por qué volver a llevarse a cabo. No tiene por qué haber una representación *posterior* donde los rasgos de alto nivel «se presenten» en una secuencia de tiempo real en beneficio de un segundo mecanismo encargado de juzgar la secuencia en cuestión. En otras palabras, una vez el cerebro ha llevado a cabo unas inferencias determinadas a partir de estas yuxtaposiciones de información temporal, éste puede representar los resultados con el formato que mejor se adecue a sus necesidades y capacidades, y no necesariamente con un formato en el que «el tiempo se utiliza para representar el tiempo».

3. Libet y el caso de la «referencia hacia atrás en el tiempo»

Hemos establecido un método mediante el cual el cerebro puede llevar a cabo sus tareas compilatorias sobre la información temporal de una manera que ignora la organización temporal real (el «tiempo de llegada») de algunas de sus representaciones, pero debemos recordar una vez más la presión temporal bajo la cual tiene que hacer todo esto. Si procedemos hacia atrás desde el punto en que finalizó el plazo, todo el contenido referido o que ha sido expresado en el comportamiento ulterior debe haber estado presente (en el cerebro, pero no necesariamente «en la conciencia») a tiempo para contribuir causalmente a dicho comportamiento. Por ejemplo, si durante un experimento un sujeto *dice* «perro» como respuesta a un estímulo visual, podemos proceder hacia atrás desde el instante en que se produjo esta conducta, que estaba claramente controlada por un proceso que tenía el contenido *perro* (a menos que el sujeto diga «perro» ante cualquier estímulo, o se pase el día diciendo «perro perro perro...», etc.). Habida cuenta de que se necesitan del orden de los 100 mseg para *empezar* a ejecutar una intención de habla de este tipo (y, aproximadamente, otros 200 mseg para completarla), podemos estar seguros de que el contenido *perro* ya estaba presente en las áreas del lenguaje del cerebro (o en su proximidad) unos 100 mseg antes de que la proferencia empezara. Si partimos del otro extremo de la cadena, podemos determinar el instante más temprano en que el contenido *perro* puede haber sido computado o extraído por el sistema visual a partir de la entrada retiniana e incluso, quizá, seguir su creación y su trayectoria posterior a través del sistema visual y hacia las áreas del lenguaje.

Lo realmente anómalo (que provocaría lamentaciones y chirriar de dientes) sería que el tiempo que pasara entre el estímulo *perro* y la proferencia de «perro» fuera menor que el tiempo físicamente necesario para el establecimiento de ese contenido y su desplazamiento por el sistema. Pero tales anomalías nunca se han detectado. Sin embargo, sí se han descubierto unas yuxtaposiciones sorprendentes entre las dos secuencias representadas en la figura 5.12 de la página 150. Cuando intentamos poner la secuencia de even-

tos del flujo objetivo de procesamiento en el cerebro de acuerdo con la secuencia objetiva del sujeto *determinada a partir de lo que con posterioridad dice el propio sujeto,* a veces encontramos unos bucles sorprendentemente grandes. Ésta es, por lo menos, la conclusión a la que querríamos llegar a la vista de uno de los experimentos más discutidos —y criticados— de la neurociencia: el experimento de neurocirugía de Benjamin Libet, que demuestra lo que él mismo ha denominado «referencia hacia atrás en el tiempo».

En ocasiones, durante la cirugía cerebral, es importante que el paciente esté despierto y alerta, únicamente bajo los efectos de una anestesia local (como cuando el dentista nos inyecta novocaína). Ello permite al neurocirujano obtener testimonios inmediatos sobre lo que experimenta el paciente cuando se sondea su cerebro (véase la nota 5 de la página 70). El pionero de esta práctica fue Wilder Penfield (1958), y durante más de treinta años los neurocirujanos han podido ir recogiendo datos sobre los efectos de la estimulación eléctrica directa de varias partes del córtex. Desde hace tiempo se sabe que la estimulación de ciertos puntos en la corteza *somatosensorial* (una franja fácilmente accesible situada en la parte superior del cerebro) produce en el paciente la experiencia de sensaciones en las partes del cuerpo correspondientes. Por ejemplo, la estimulación de un punto en la parte izquierda de la corteza somatosensorial puede producir la sensación de un breve hormigueo en la mano derecha del sujeto (con motivo de la conocida inversión en el sistema nervioso que hace responsable de la parte derecha del cuerpo a la mitad izquierda del cerebro, y viceversa). Libet comparó los tiempos de dichos hormigueos inducidos por estimulación del córtex con los de sensaciones similares provocadas de la manera usual, como por ejemplo la aplicación de un impulso eléctrico directamente sobre la mano (Libet, 1965, 1981, 1982, 1985b; Libet y otros, 1979; véase también Popper y Eccles, 1977; Dennett, 1979b; Churchland, 1981a, 1981b; Honderich, 1984).

¿Qué debemos esperar que ocurra? Imaginemos a dos personas que se dirigen a su trabajo cada día a la misma hora, con la única diferencia de que uno de ellos vive en las afueras de la ciudad y el otro sólo a unas manzanas del puesto de trabajo. Ambos conducen a la misma velocidad, así que, dada la distancia extra que debe recorrer el primero, es de esperar que llegue a la oficina más tarde que el segundo. Sin embargo, no es esto lo que halló Libet cuando le preguntó a sus pacientes qué pasaba primero, si el hormigueo en la mano que se iniciaba directamente en el córtex o el hormigueo enviado desde la mano. A partir de los datos obtenidos argumentó que, aunque en cada caso había un espacio de tiempo considerable (500 mseg aproximadamente) entre el principio de la estimulación y la «adecuación neuronal» (el punto en que, según Libet, culminan los procesos corticales a fin de dar lugar a la experiencia consciente del hormigueo), cuando se estimulaba la mano, la experiencia se «refería hacia atrás en el tiempo» «automáticamente», y se percibía *antes* que el hormigueo producido por la estimulación del cerebro.

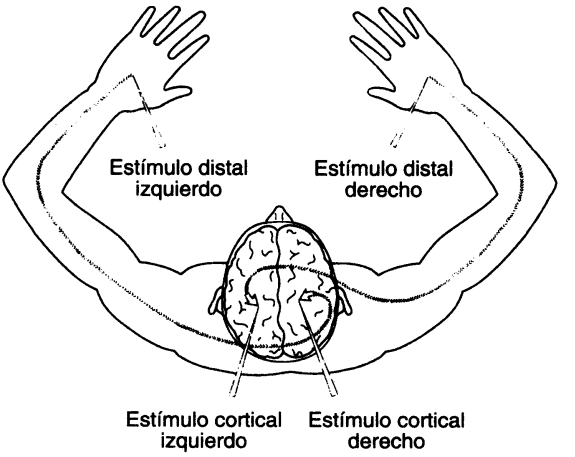


Figura 6.2

Aún más sorprendente, Libet relata casos en los que la parte izquierda del *córtex* del paciente era estimulada antes que su *mano* izquierda, situación en la que todos tendemos a pensar que se percibirían dos hormigueos: primero en la mano derecha (el inducido corticalmente) y después en la mano izquierda. El testimonio subjetivo era, no obstante, el inverso: «primero izquierda, después derecha».

Libet ha interpretado sus resultados como un serio desafío para el materialismo: «...una disociación entre la organización temporal de los eventos "mentales" y "físicos" correspondientes plantearía serias dificultades, aunque no dificultades insuperables, a la... teoría de la identidad psiconeuronal» (Libet y otros, 1979, pág. 222). Según sir John Eccles, Premio Nobel de medicina por sus investigaciones en neurofisiología, no se puede hacer frente a dicho desafío:

> Este procedimiento de anticipación no parece explicable mediante ningún proceso neurofisiológico. Es de presumir que se trate de una estrategia aprendida por la mente autoconsciente... [E]l adelantamiento de la experiencia sensorial es atribuible a la capacidad de la mente autoconsciente de realizar pequeños ajustes temporales, esto es, de hacer trampas con el tiempo (Popper y Eccles, 1977, pág. 364).*

Más recientemente, el físico y matemático Roger Penrose (1989) ha sugerido que una explicación materialista del fenómeno descrito por Libet exigiría que se llevara a cabo una revolución en la física fundamental. Aunque

* Página 409 de la edición española. [N. del T.]

el experimento de Libet se ha esgrimido frecuentemente en círculos no científicos como una demostración de la veracidad del dualismo, muy pocos hoy en día dentro de la comunidad de la ciencia cognitiva comparten esta opinión. En primer lugar, los procedimientos experimentales de Libet y su análisis de los resultados, han recibido críticas muy severas. Su experimento nunca ha sido replicado, lo cual es para muchos motivo suficiente para eliminar toda consideración de sus «resultados». La visión escéptica consiste, pues, en suponer que estos fenómenos simplemente no existen. *Pero, ¿y si existieran?* Éste es precisamente el tipo de pregunta que haría un filósofo, aunque, en este caso, hay algo más que meras motivaciones filosóficas para hacerla. Nadie duda de la existencia de fenómenos simples tales como el fenómeno phi de los colores o el del conejo cutáneo, cuya interpretación plantea problemas similares. El apelar exclusivamente a motivos de orden *metodológico* para desechar los resultados de Libet sería un acto flagrante de miopía teórica, en particular porque esta actitud no cuestiona los supuestos básicos que parecen indicar que si el experimento de Libet *llegara* algún día a ser replicado adecuadamente, entonces se avecinarían días muy sombríos para el materialismo.

Lo primero que merece la pena observarse sobre el experimento de Libet es que éste no constituiría evidencia de anomalía alguna si renunciáramos a la oportunidad de registrar los testimonios orales de las experiencias de los sujetos y de utilizarlos para generar un texto primero y un mundo heterofenomenológico después. Los sonidos que emiten con su tracto vocal durante o después de los experimentos no indican la presencia de paradoja alguna si los tratamos exclusivamente como fenómenos acústicos. En ningún caso, los sonidos parecen salir de las cabezas sin que se hayan movido los labios primero ni las manos se mueven antes que los eventos cerebrales que supuestamente causan este movimiento se hayan producido, ni tampoco se producen eventos en el córtex antes que los estímulos que los originan. Vistos estrictamente como conductas internas y externas de un sistema de control para un cuerpo implementado sobre un soporte biológico, los acontecimientos observados y registrados en los experimentos no constituyen una muestra aparente de violación alguna de la causalidad mecánica tradicional, es decir, aquella cuyo modelo estándar aproximado sería el de la física galileana/newtoniana.

Usted podría «hacer desaparecer los fenómenos» si fuera un defensor del conductismo desnudo y rechazara tomar en serio los testimonios. Pero nosotros no aceptamos el conductismo desnudo; queremos aceptar el desafío de otorgar un sentido a lo que Libet denomina «un aspecto fenomenológico central de la existencia humana en relación a la función cerebral» (1985a, pág. 534). Libet se aproxima mucho a la heterofenomenología. Escribe: «Es importante darse cuenta de que estas referencias y correcciones subjetivas tienen lugar aparentemente en el plano de la "esfera" *mental;* no se manifiestan como tales en las actividades de los niveles neuronales» (1982, pág. 241). Pero, ya que no dispone de un método *neutral* para hacer referencia

a la fenomenología, se ve obligado a asignar la anomalía al plano de «la "esfera" mental». Un pequeño paso, forzado (porque *tiene* que darlo si quiere rechazar el conductismo), pero un primer paso sobre la rampa deslizante que devuelve al dualismo.

> Los testimonios de los sujetos sobre sus diversas experiencias... no eran constructos teóricos, sino observaciones empíricas. (...) Puede que el método introspectivo tenga sus limitaciones, pero puede utilizarse con propiedad dentro del marco de las ciencias naturales y es, además, esencial si estamos intentando obtener datos experimentales sobre el problema de la mente y el cerebro (1987, pág. 785).

Los testimonios de los sujetos, incluso una vez se los ha convertido en textos, son, como señala Libet, observaciones empíricas, pero *lo que los sujetos refieren en estos testimonios,* los eventos en sus mundos heterofenomenológicos, sí que son constructos teóricos. Según nos recomienda encarecidamente Libet, pueden ser utilizados con propiedad dentro del marco de las ciencias naturales, pero *sólo* si se interpretan desde el principio como las ficciones de un teórico.

Libet afirma que sus experimentos sobre la estimulación directa del córtex ponen de manifiesto «dos factores temporales fundamentales»:

1. *Existe un retraso sustancial antes del instante en que las actividades cerebrales,* iniciadas por un estímulo sensorial, *alcanzan la «adecuación neuronal»* a fin de provocar cualquier experiencia sensorial consciente como resultado.
2. Una vez se ha alcanzado la adecuación neuronal, la organización temporal subjetiva de la experiencia *se refiere hacia atrás en el tiempo (automáticamente),* utilizando una «señal de organización temporal» en forma de respuesta inicial del córtex cerebral al estímulo sensorial (1981, pág. 182).

La «señal de organización temporal» es la *primera* muestra de actividad en el córtex (el potencial evocado primario), que se produce solamente unos 10 o 20 mseg después de la estimulación del órgano sensorial periférico. Libet sugiere que la referencia hacia atrás es siempre referencia «a» esa señal de organización temporal.

El modelo de Libet es estaliniano: después del potencial evocado primario, se produce en el córtex toda una serie de procesos de edición antes del instante en que se alcanza la «adecuación neuronal», momento en el que se proyecta la película acabada. ¿Cómo se proyecta? Aquí, Libet duda entre una visión radical y otra visión más moderada (véase Honderich, 1984):

1. *Proyección hacia atrás:* la película se envía hacia atrás en el tiempo a un Teatro Cartesiano donde se proyecta en sincronía con los poten-

ciales evocados primarios. (Los potenciales evocados primarios, en tanto que «señales de organización temporal», cumplen una función parecida a la de la claqueta en el momento de rodar una película, mostrando al proyector exactamente a qué punto hacia atrás en el tiempo debe enviarse la experiencia.)

2. *Referencia hacia atrás:* la película se proyecta en el tiempo ordinario, pero lleva algo parecido a un matasellos, recordando así al espectador que esos acontecimientos deben ser interpretados como eventos que se produjeron en un tiempo anterior. (En este caso, los potenciales evocados primarios funcionan simplemente como fechas, que podrían *representarse* en la pantalla cartesiana mediante títulos como «En la víspera de la batalla de Waterloo» o «Ciudad de Nueva York, verano de 1942».)

Libet utiliza el término *referencia,* y lo defiende recordándonos el fenómeno «ampliamente reconocido y aceptado» de la referencia espacial, lo que parece apuntar hacia la versión moderada.

> El de la referencia subjetiva hacia atrás en el tiempo es un concepto extraño y quizá difícil de aceptar la primera vez que uno se topa con él. Sin embargo, existe un claro precedente en el ampliamente reconocido y aceptado concepto de la referencia subjetiva en la dimensión espacial. Por ejemplo, la imagen visual experimentada como respuesta a un estímulo visual posee una configuración y una localización espaciales que son muy distintas de la configuración y localización espaciales de las actividades neuronales que dan lugar a la imagen («subjetivamente referida»). (1981, pág. 183. Véase también Libet y otros, 1979, pág. 221; Libet, 1985b.)

Sin embargo, su conclusión es que la referencia temporal supone un problema para el materialismo (la «teoría de la identidad psiconeuronal»: Libet y otros, 1979, pág. 222), así que o cree que la referencia espacial también plantea los mismos problemas, o no ha comprendido la defensa de sus propias ideas. Porque si la referencia espacial —el hecho de que lo que vemos parece estar fuera, y no dentro, de nuestros cerebros— supone un problema para el materialismo, entonces ¿por qué Libet sugiere que sus trabajos aportan un nuevo e importante argumento en favor del dualismo? No cabe duda de que el fenómeno de la referencia espacial es más conocido que el tipo de referencias temporales que Libet pone tanto empeño en demostrar. Parece, sin embargo, que tiene una visión radical (o cuando menos confusa) de la referencia espacial en tanto que una especie de «proyección»:

> Existen datos experimentales que avalan la visión de que la «esfera» mental o subjetiva podría «rellenar» lagunas espaciales y temporales. ¿Cómo, si no, podemos interpretar, por ejemplo, la notable discrepancia, mencionada anteriormente, *que sabemos que existe* entre una imagen visual subjetiva y la configuración de actividades neuronales que da lugar a la experiencia de la imagen en cuestión? (1981, pág. 196).

Ello parece indicar que el proyector que Smythies no pudo encontrar en el cerebro se halla, de hecho, oculto en la «esfera» mental.[8]

¿De qué manera afirma Libet poder establecer sus dos sorprendentes factores temporales? La «adecuación neuronal», que según Libet precisa unos 500 mseg de actividad cortical, se determina comprobando el tiempo que una estimulación cortical directa puede tardar, después de la estimulación inicial, en *interferir con* la conciencia atestiguada con posterioridad. Más allá del intervalo crítico, un estímulo cortical directo sería referido por el sujeto como una experiencia *subsiguiente.* (Al haber llegado demasiado tarde para ser incorporado por la sala de edición en la «impresión final» de la experiencia del primer estímulo, aparecerá en la siguiente entrega.) Los datos de Libet sugieren la presencia de una ventana de edición tremendamente variable: «El estímulo cortical condicionante podía iniciarse más de 500 mseg después del impulso sobre la piel y aun modificar la sensación de la piel, aunque en la mayoría de los casos no se observaron efectos retroactivos con intervalos S-C mayores de 200 mseg» (1981, pág. 185). Libet tiene el cuidado de definir la adecuación neuronal en términos del efecto sobre los testimonios orales producidos después de un breve tiempo de reflexión: «Se pedía al sujeto que relatara sus experiencias unos segundos después de la producción de cada par de... estímulos» (1979, pág. 195), e insiste en que «la organización temporal de la experiencia subjetiva debe distinguirse de la organización temporal de la respuesta comportamental (como en el tiempo de reacción), la cual puede producirse antes de que el conocimiento consciente se haya desarrollado...» (Libet y otros, 1979, pág. 193).

Esta salvedad le permite defenderse de una interpretación rival de Patricia Churchland basada en datos propios. Churchland es la primera «neurofilósofa» (véase su libro de 1986, *Neurophilosophy: Toward a Unified Science of the Mind-Brain*). Cuando leí por primera vez los resultados de Libet (en Popper y Eccles, 1977), la animé a estudiarlos, y ella les propinó una vigorosa sacudida (Churchland, 1981a). Intentó desacreditar la primera tesis de Libet, el largo espacio de tiempo necesario para alcanzar la «adecuación neuronal» para la conciencia, pidiendo a los sujetos que dijeran *go* tan pronto

8. Véase también el rechazo que expresa Libet ante la sugerencia de MacKay para una lectura más moderada (1981, pág. 195; 1985b, pág. 568). Por otra parte, su conclusión final en 1981 es todo menos concluyente: «Mi visión personal... ha sido la de que la discrepancia temporal crea ciertas dificultades para la teoría de la identidad, pero que estas dificultades no son insuperables» (pág. 196). Presumiblemente serían totalmente insuperables en la interpretación que comporta una *proyección* hacia atrás, ya que ésta presupone precognición o causalidad retroactiva u otro mecanismo igualmente extraño y desconocido. Además, más tarde Libet (1985b, pág. 569) describe estas dificultades superables de un modo que parece invitar a una lectura más moderada: «Aunque la hipótesis del retraso-más-anticipación no separa el tiempo real de la experiencia del tiempo de su producción neuronal, no elimina sin embargo la necesidad de apelar a la simultaneidad entre *la organización temporal subjetiva* de la experiencia y el tiempo real objetivo de la experiencia». Es posible que la vehemente defensa que sir John Eccles ha llevado a cabo de la interpretación radical y dualista de estos hallazgos haya desviado la atención de Libet (y de sus críticos) de las tesis moderadas que él mismo a veces mantiene.

como fuesen conscientes de un estímulo sobre la piel como los que usó Libet. Con nueve sujetos, obtuvo un tiempo medio de respuesta de 358 mseg, lo cual, según ella, demostraba que los sujetos debían haber alcanzado la adecuación neuronal en la cota de los 200 mseg como muy tarde (dejando tiempo para la producción de la respuesta verbal).

La respuesta de Libet es estaliniana: una reacción verbal —como decir *go*— puede iniciarse inconscientemente. «Nada hay de mágico o particularmente informativo cuando la respuesta motora es una vocalización de la palabra *go* en lugar de la más usual del dedo presionando un botón. (...) La capacidad de detectar un estímulo y de reaccionar a él intencionadamente, o de estar influido psicológicamente por él, sin que por ello haya un conocimiento consciente del estímulo que pueda ser referido, es un hecho ampliamente aceptado» (1981, págs. 187-188). Y ante la objeción siguiente: «¿Y qué creían los sujetos de Churchland que estaban haciendo si no decir, como se les había pedido, en qué momento eran conscientes del estímulo?», Libet podía dar la respuesta estaliniana estándar: Sí, éstos llegaron finalmente a ser conscientes del estímulo, pero para entonces su testimonio oral ya se había iniciado.[9]

Por este motivo, Libet rechaza aquellos estudios de tiempos de reacción como los de Churchland, porque tienen «una validez dudosa como criterio básico para la experiencia consciente» (1981, pág. 188). Él prefiere que el sujeto se tome su tiempo: «El testimonio se lleva a cabo sin prisas unos segundos después de cada intento, lo cual permite al sujeto examinar su evidencia introspectivamente» (pág. 188). ¿Cómo trata, pues, la idea alternativa de que toda esta parsimonia da al revisionista orwelliano tiempo suficiente en el cerebro para sustituir los recuerdos *reales* de la conciencia por recuerdos *falsos*?

> Evidentemente, el testimonio después del intento requiere que los procesos de la memoria a corto plazo y del recuerdo sean operativos, pero ello no presenta dificultad alguna para los sujetos cuyas capacidades no sufren de defectos significativos (pág. 188).

Ello plantea una petición de principio en contra del orwelliano, quien puede explicar toda una serie de efectos como el resultado de errores *normales* en el recuerdo o de un recuerdo alucinatorio, en el cual un evento real anterior de la conciencia es eliminado y sustituido por recuerdos posteriores.

9. En un artículo anterior, Libet considera la posibilidad de procesos orwellianos y supone que podría haber una diferencia significativa entre eventos mentales inconscientes y eventos mentales conscientes pero efímeros: «Podría muy bien ser que hubiera un tipo inmediato pero efímero de experiencia consciente que no es retenida para el recuerdo en los niveles conscientes de la experiencia. Sin embargo, si tales experiencias existen, su contenido deberá tener una significación directa tan sólo en procesos mentales inconscientes tardíos, aunque, al igual que otras experiencias inconscientes, podrían jugar un papel indirecto en experiencias conscientes posteriores» (1965, pág. 78).

¿Acaso Libet ha dejado cocer el guiso demasiado tiempo, o quizás es que Churchland lo ha probado demasiado pronto? Si Libet quiere defender su idea de que su elección del tiempo de sondeo tiene un estatuto *privilegiado,* entonces debe estar preparado para combatir los argumentos en contra.

Libet está a punto de alegar *nolo contendere:* «Es preciso admitir que un testimonio del orden temporal relativo no puede, por sí mismo, ser un indicador del tiempo "absoluto" (tiempo de reloj) de la experiencia: como ya hemos sugerido, no existe método conocido para establecer dicho indicador» (1981, pág. 188). Esto nos trae a la memoria una afirmación anterior según la cual no parece haber «un método a partir del que uno pueda determinar la organización temporal absoluta de una experiencia subjetiva» (Libet y otros, 1979, pág. 193). Libet no acierta a ver, sin embargo, la posibilidad de que ello se deba al hecho de que dicho momento de tiempo absoluto no existe (véase Harnad, 1989).

En sus críticas, Churchland (1981a, 1981b) también comete el error de no distinguir el tiempo representado, del tiempo de representación:

> Las dos hipótesis se diferencian esencialmente por el momento en que *se perciben* las sensaciones respectivas (la cursiva es mía; 1981a, pág. 177).

> Aun si supusiéramos que las sensaciones que surgen de sensaciones simultáneas en la piel y en el LM *(lemniscus medialis) se percibieran exactamente al mismo tiempo,* el retraso en la adecuación neuronal para el estímulo en la piel podría muy bien ser un efecto creado por la propia organización (1981b, pág. 494).

Supongamos que se eliminaran todos estos efectos y que, aun así, las sensaciones «se percibieran exactamente al mismo tiempo». ¿Cómo interpretaría Churchland este resultado inesperado? ¿Significaría que hay un tiempo *t* tal que el estímulo 1 se percibe en *t* y el estímulo 2 se percibe en *t* (el panorama antimaterialista) o solamente que el estímulo 1 y el estímulo 2 se perciben (se experimentan) como simultáneos? Churchland no consigue evitar que pensemos que los hallazgos de Libet, si llegaran a confirmarse, causarían estragos (como él mismo ha afirmado alguna vez) en el materialismo. En otra parte, sin embargo, señala correctamente que «por muy intrigantes que resulten las ilusiones temporales, no hay motivo para suponer que éstas tengan nada de preternatural, y, sin duda, no hay nada que las distinga de las ilusiones espaciales o las ilusiones con el movimiento como las únicas portadoras de una denominación de origen no física» (1981a, pág. 179). Esto sólo podría ser si las ilusiones temporales fuesen fenómenos en los que *el tiempo se representa mal;* si las *malas representaciones* tienen lugar en los momentos «equivocados», entonces es que se está tramando algo más revolucionario.

¿En qué lugar quedan los experimentos de Libet con la estimulación cortical? Como un intento interesante pero inconcluyente de establecer el modo

en que *el cerebro representa el orden temporal.* Los potenciales evocados primarios podrían servir de puntos de referencia específicos para las representaciones neuronales del tiempo, aunque Libet no lo ha demostrado, como queda claro en las críticas más técnicas de Churchland. Alternativamente, el cerebro podría hacer uso de unas representaciones del tiempo más lábiles. No representamos los objetos que vemos tal y como aparecen en la retina, sino a diferentes distancias en el mundo exterior; ¿por qué no podría el cerebro representar los eventos produciéndose *en el momento* en que existen más motivos «ecológicos» para que se produzcan? Cuando estamos ocupados en una tarea de habilidad manual, el tiempo estándar debería ser el «tiempo de la punta de los dedos»; cuando estamos dirigiendo una orquesta, el «tiempo del oído» podría captar la grabación. El «tiempo cortical primario» podría ser el estándar por defecto (algo parecido a la hora de Greenwich para el imperio británico), aunque esto es tema de ulteriores investigaciones.

La cuestión ha quedado oscurecida por el hecho de que el proponente y el crítico no han sido consistentes en el momento de distinguir el tiempo del representante del tiempo del representado. Es un diálogo de sordos, con Libet defendiendo su posición estaliniana y Churchland haciendo sus contraataques orwellianos, ambos aparentemente de acuerdo en que hay un hecho verdadero sobre cuándo exactamente (en tiempo «absoluto», como diría Libet), se produce una experiencia consciente.[10]

4. Las afirmaciones de Libet sobre el retraso subjetivo de la conciencia de las intenciones

Libet explota el concepto de organización temporal absoluta de una experiencia en una serie de experimentos posteriores con «intenciones conscientes». En estos experimentos, intentaba determinar experimentalmente esta organización temporal absoluta, permitiendo a los sujetos, quienes por sí solos tienen (en cierto modo) acceso directo a sus experiencias, hacer *su autoorganización temporal.* Solicitó a unos sujetos normales (no sometidos a neurocirugía) que efectuaran decisiones «espontáneas» de flexionar una mano por la muñeca mientras anotaban la posición de un punto en un disco giratorio (la «segunda manecilla» de un reloj, de hecho) en el momento preciso en que se formaban la intención de efectuar la flexión (Libet, 1985a, 1987, 1989). Después (unos segundos más tarde), los sujetos debían señalar dónde estaba el punto en el momento en que decidieron flexionar la muñeca. Esto

10. Harnad (1989) percibe un problema de medición irresoluble, pero niega precisamente lo que yo estoy afirmando: que no hay tal hecho verdadero: «La introspección sólo puede decirnos en qué momento *pareció* producirse un evento determinado, o cuál de los dos eventos *pareció* producirse el primero. No hay un modo independiente de confirmar que la organización temporal real fuese realmente como pareció que era. La inconmensurabilidad es un problema metodológico, no metafísico» (pág. 183).

permitió a Libet calcular el instante exacto (con una precisión de milisegundos) en que los sujetos *pensaron* que habían tomado la decisión, y comparar este instante de tiempo con la organización temporal de los eventos concurrentes que se producían en el cerebro de los sujetos. Halló así datos que evidenciaban un retraso de estas «decisiones conscientes» situado entre los 350 y 400 mseg después del inicio de los «potenciales de disponibilidad» que pudo registrar a través de unos electrodos en el cuero cabelludo que, según él, captaban los eventos neuronales que determinan las acciones voluntarias llevadas a cabo. Su conclusión es que «la iniciación cerebral de un acto voluntario espontáneo empieza de manera inconsciente» (1985a, pág. 529).

Ello parece demostrar que la conciencia va rezagada con respecto a los procesos cerebrales que realmente controlan el cuerpo. Para muchos, éste es un panorama inquietante e incluso deprimente, ya que parece negar todo «poder ejecutivo» real (al contrario que ilusorio) al «yo consciente». (Véanse las discusiones sobre este asunto por parte de los comentaristas de Libet, 1985a, 1987, 1989; véanse también Harnad, 1982; Pagels, 1988, pág. 233 y sigs.; y Calvin, 1989a, págs. 80-81.)

Libet es mucho más claro que algunos de sus críticos sobre la importancia de mantener la distinción entre contenido y vehículo: «No debe confundirse *lo que* refiere el sujeto con *cuándo* éste puede llegar a ser introspectivamente consciente de lo que está refiriendo» (Libet, 1985a, pág. 559). Reconoce (pág. 560), asimismo, que un juicio de simultaneidad no tiene necesariamente por qué producirse ni representarse simultáneamente; puede madurar durante un cierto período de tiempo (considérese, por ejemplo, el tiempo que pueden necesitar los jueces de una carrera para revelar y después examinar la *foto-finish* con la que finalmente emiten su juicio sobre quién es el ganador o sobre si hay un empate).

Libet recogió datos sobre dos series temporales:

> la serie objetiva, que incluye la organización temporal del reloj externo y de los eventos neuronales relevantes: los potenciales de disponibilidad (PD) y los electromiogramas (EMG) que registraban el inicio de la contracción del músculo;
>
> la serie subjetiva (referida con posterioridad), que consiste en imágenes mentales, en recuerdos de cualquier acto previo de planificación, y en un único dato de referencia para cada intento: un juicio de simultaneidad cuya forma general sería *mi intención consciente (W) empezó simultáneamente con el punto del reloj en la posición P.*

Libet parece haber querido hallar un equivalente aproximado del esquivo *acte gratuit* de que hablaban los existencialistas (por ejemplo Gide, 1948; Sartre, 1943), la pura elección carente de motivo y, por tanto y en cierto sentido, «libre». Como muchos comentaristas han señalado, estas acciones tan poco usuales (que podrían denominarse actos de deliberada pseudoaleatoriedad) apenas pueden ser considerados como paradigmas del «acto volun-

tario normal» (Libet, 1987, pág. 784). Pero ¿consiguió Libet aislar en algún momento un determinado tipo de experiencias conscientes, como sea que las caractericemos, a las que es posible atribuir una organización temporal absoluta a partir del diseño experimental descrito?

Libet sostiene que cuando las intenciones conscientes de llevar a cabo un acto (cuando menos aquellas que son de ese tipo especial que él describe) se hacen encajar con los eventos mentales que realmente inician los actos, hay un tiempo de compensación que oscila entre los 300 y los 500 mseg. Esto es mucho tiempo —casi medio segundo—, y una amenaza para todos los que están comprometidos con el principio de que nuestros actos conscientes *controlan* nuestros movimientos corporales. Es como si *nosotros* estuviéramos en el Teatro Cartesiano donde se nos muestra, con un retraso en la cinta de medio segundo, el acto *real* de toma de decisiones que tiene lugar *en otra parte* (un lugar en el que *nosotros* no estamos). No es que «no estemos en la onda» (como dicen en la Casa Blanca), pero ya que accedemos a la información con retraso, lo único que podemos hacer es intervenir con «vetos» o «desencadenamientos» de última hora. Corriente abajo desde el cuartel general (consciente), yo ya tengo iniciativa real, nunca estoy presente en el nacimiento de un proyecto, pero ejerzo una moderada modulación ejecutiva de las políticas previamente dictadas que vienen a parar a mi despacho.

Esta imagen es atractiva, pero incoherente. Este modelo de Libet, como el anterior, es estaliniano y, por tanto, tiene una alternativa orwelliana: los sujetos *fueron* conscientes de sus intenciones en un momento temprano, pero esta conciencia fue borrada de la memoria (o simplemente revisada) antes de que tuvieran oportunidad de recordarla. Libet admite que este hecho «plantea un problema, pero no se puede evaluar experimentalmente» (1985a, pág. 560).

Dada esta concesión que hace Libet, ¿está mal concebida entonces la tarea de fijar la organización temporal absoluta de la conciencia? Ni Libet ni sus críticos llegan a tal conclusión. Éste, a pesar de haber distinguido claramente entre vehículo y contenido —*lo que* se representa y el *momento* en que se representa—, intenta inferir, a partir de premisas sobre lo que se representa, conclusiones sobre la organización temporal absoluta del representante en la conciencia. El psicólogo Gerald Wasserman (1985, pág. 556) ha sabido ver el problema: «Es posible determinar con facilidad el instante en que el punto objetivo extremo ocupa una determinada posición en el reloj, pero no es éste el resultado que buscamos». Pero finalmente cae también en la trampa cartesiana: «Lo que necesitamos es el tiempo en que se produce la representación interna del punto en la mente-cerebro».

¿«*El* tiempo en que se produce» la representación interna? ¿Dónde se produce? Esencialmente, lo que hay es una representación continua del punto (representándolo en posiciones distintas) en diferentes partes del cerebro, empezando por la retina y siguiendo hacia arriba a lo largo del sistema visual. El brillo del punto se representa en algunas partes y en algunos tiem-

pos, su localización en otros y su movimiento otros más. A medida que el punto externo se mueve, todas estas representaciones cambian de una manera asincrónica y espacialmente distribuida. ¿Dónde «se reúne todo en un instante de la conciencia»? En ninguna parte.

Wasserman acierta al señalar que la tarea del sujeto de determinar dónde estaba el punto en un instante de la secuencia subjetiva es, en sí misma, una tarea voluntaria, e iniciarla seguramente requiere su tiempo. Esto es difícil no sólo porque compite con otros proyectos simultáneos, sino también porque no es natural: un juicio consciente de temporalidad de un tipo que normalmente no juega ningún papel en el control de la conducta y que, por tanto, carece de significado natural en la secuencia. El proceso de interpretación que finalmente fija el juicio de simultaneidad subjetiva es también un producto de la situación experimental y, además, *altera el objetivo de la tarea*, motivo por el cual no nos dice nada interesante sobre la organización temporal real de los vehículos normales de representación del cerebro.

La visión demasiado natural que debemos rechazar es la siguiente: en algún lugar de las profundidades del cerebro empieza la iniciación de un acto; comienza como una intención inconsciente y, poco a poco, va avanzando hacia el Teatro, ganando en precisión y poder a medida que avanza, y entonces, en un instante *t*, sale a escena, donde se lleva a cabo el desfile de las representaciones de los puntos, llegadas desde la retina después de ser revestidas de atributos como el brillo y la localización, que recibieron a medida que se movían. La audiencia, o yo, tiene la tarea de decir qué representación del punto estaba «en escena» exactamente en el momento en que la intención consciente hizo su entrada. Una vez identificado, el tiempo de partida desde la retina de este punto puede ser calculado, así como la distancia que lo separa del Teatro y la velocidad de transmisión. Es así que podemos determinar el momento exacto en que la intención consciente se produjo en el Teatro Cartesiano.

No entiendo por qué esta idea resulta tan seductora. ¡Es tan fácil de visualizar! ¡Parece tan apropiada! ¿Acaso no es esto lo que *tiene* que ocurrir cuando dos cosas tienen lugar a la vez en la conciencia? No. De hecho, esto *no puede* ser lo que ocurre cuando dos cosas tienen lugar a la vez en la conciencia, porque en el cerebro no existe ningún lugar como éste. Algunos han pensado que la incoherencia de *esta* concepción no obliga a abandonar la idea de una organización temporal absoluta de las experiencias. Aparentemente, existe un modelo alternativo para el inicio de la conciencia que evita la absurdidad cartesiana de un centro en el cerebro sin por ello impedir la organización temporal absoluta. ¿No podría ser que la conciencia fuera una cuestión, no de llegada a un punto determinado, sino de que una representación supere un determinado umbral de activación en todo el córtex o en gran parte del mismo? De acuerdo con este modelo, un elemento de contenido se hace consciente en un instante *t*, no por el hecho de entrar en un sistema funcionalmente definido y anatómicamente localizado, sino por el hecho de cambiar de estado allí donde se encuentra: adquiriendo alguna propiedad

o viendo la intensidad de alguna de sus propiedades potenciada por encima de un umbral determinado.

La idea de que la conciencia es un *modo de acción* del cerebro en vez de un *subsistema* del mismo es muy recomendable (véase, por ejemplo, Kinsbourne, 1980; Neumann, 1990; Crick y Koch, 1990). Además, es probable que dichos cambios de modo puedan ser medidos por observadores exteriores, lo que proporcionaría, en principio, una secuencia de contenidos única y determinada adquiriendo un modo especial. Sin embargo, esto sigue siendo el Teatro Cartesiano si lo que se afirma es que la organización temporal real («absoluta») de dichos cambios de modo es lo que define la secuencia subjetiva. La imagen es ligeramente distinta, pero las implicaciones son las mismas. El hecho de conferir una propiedad especial que caracteriza la conciencia en un instante es sólo parte del problema; discriminar que la propiedad ha sido conferida es la otra, y aunque observadores científicos con sus instrumentos sean capaces de hacerlo con una precisión de microsegundos, ¿cómo lo hace el cerebro?

Nosotros los seres humanos emitimos juicios de simultaneidad y de secuencia de los elementos de nuestra propia experiencia, algunos de los cuales expresamos abiertamente, de modo que en algún punto o puntos de nuestro cerebro se cruza la barrera que separa la organización temporal de las representaciones de la representación de la organización temporal; dondequiera y cuandoquiera que se lleven a cabo dichas discriminaciones, después, las propiedades temporales de las representaciones que encarnan a estos juicios no son constitutivas de su contenido. Las simultaneidades objetivas y las secuencias de eventos que se distribuyen por la amplia superficie del córtex carecen de relevancia funcional *a menos que también puedan ser detectadas con precisión por los mecanismos cerebrales.* La clave del asunto puede formularse en forma de pregunta: ¿qué hace que *esta* secuencia se convierta en el flujo de la conciencia? No hay nadie ahí dentro, *mirando* la pantalla gigante que retransmite el espectáculo por todo el córtex, incluso si el espectáculo es discernible para los observadores *exteriores.* Lo que importa es la manera en que esos contenidos son utilizados por o en los procesos de control de la conducta, lo cual sólo *puede* estar constreñido indirectamente por la organización temporal cortical. Lo importante, de nuevo, no son las propiedades temporales de los representantes, sino las propiedades temporales *representadas,* lo que se determina por la manera en que son «tomadas» por los procesos cerebrales subsiguientes.

5. Un regalo: el carrusel precognitivo de Grey Walter

Después de lidiar con los casos más complicados, nos merecemos un encuentro con algo extraño pero *relativamente fácil* de comprender, algo que nos devuelva al mensaje de este capítulo tan difícil. Como acabamos de observar, el experimento de Libet sobre la autoorganización temporal creaba

una tarea de emisión de juicios difícil y artificial, que nos ha privado de los resultados con la significación que esperábamos. Un experimento anterior, llevado a cabo por el neurocirujano británico W. Grey Walter (1963), no tenía este defecto. Grey Walter realizó su experimento con pacientes a los que implantó unos electrodos en el córtex motor. Su objetivo era el de verificar la hipótesis de que ciertas explosiones de actividad registradas son las iniciadoras de acciones intencionales. Lo organizó todo de manera que cada paciente mirara unas diapositivas proyectadas desde un proyector de carrusel. El paciente podía hacer avanzar el carrusel cuando quería presionando un botón en el mando de control. (Nótese la similitud con el experimento de Libet: la decisión era «libre», delimitada tan sólo por una subida endógena del aburrimiento, por la curiosidad de ver la siguiente diapositiva, por la distracción o cualquier otro motivo parecido.) Sin que lo supiera el paciente, sin embargo, el botón del mando no funcionaba, pues no estaba conectado al proyector. Lo que hacía avanzar las diapositivas era la señal amplificada proveniente del electrodo implantado en el cerebro del paciente.

Podría suponerse que los pacientes no notaban nada raro, pero el hecho es que se veían sorprendidos por el efecto, ya que les parecía que el proyector de diapositivas se anticipaba a sus decisiones. Relataban que justo en el momento en que estaban «a punto» de presionar el botón, pero antes de que realmente lo hubieran decidido, el proyector avanzaba la diapositiva, ¡y se encontraban presionando el botón con la preocupación de que la diapositiva iba a cambiar dos veces! De acuerdo con el testimonio de Grey Walter, el efecto era muy fuerte, pero aparentemente nunca llegó a realizar la obligada continuación del experimento: introducir un retraso variable a fin de comprobar la duración que éste debe tener para eliminar el efecto del «carrusel precognitivo».

Una diferencia importante entre los diseños experimentales de Grey Walter y Libet es que el juicio sobre el orden temporal que produce sorpresa en el experimento de Grey Walter es parte de una tarea normal de control de la conducta. En este sentido, es más parecido a los juicios de orden temporal por los cuales nuestros cerebros distinguen un movimiento de izquierda a derecha de un movimiento de derecha a izquierda, que a juicios de orden «deliberados y conscientes». En este caso, el cerebro se ha fijado el objetivo de «esperar» información visual sobre el éxito en la ejecución de su proyecto de hacer avanzar el carrusel, y esa información llega antes de lo esperado, disparando así una alarma. Esto nos podría enseñar algo importante sobre la organización temporal real de los vehículos de contenido y de los procesos concomitantes en el cerebro, pero, a pesar de las apariencias, no nos enseñará nada sobre la «organización temporal de la decisión consciente de cambiar la diapositiva».

Supóngase, por ejemplo, que una ampliación del experimento de Grey Walter demostrara que fuera un retraso de 300 mseg (como el implicado por las experiencias de Libet) el que tuviera que incorporarse en el sistema a fin de eliminar la sensación subjetiva del cambio de diapositivas precogni-

tivo. Lo que este retraso demostraría, de hecho, sería que las expectativas fijadas por una decisión de cambiar la diapositiva se sintonizan para esperar la información visual 300 mseg más tarde, y que se debe reaccionar con una alarma en otras condiciones. (Parecido a un mensaje de conmoción del comandante en jefe de Calcuta dirigido a Whitehall después de la batalla de Nueva Orleans.) El hecho de que la alarma *finalmente* sea interpretada en la secuencia subjetiva como una percepción de eventos desordenados (cambiar antes la presión en el botón) no demuestra nada sobre en *qué* instante del tiempo real ocurrió por primera vez la conciencia de la decisión de presionar el botón. La sensación que refirieron los sujetos de no tener tiempo de «vetar» el acto de presionar el botón que habían iniciado cuando «vieron que la diapositiva ya estaba cambiando» es una interpretación natural por parte del cerebro a fin de ordenar (finalmente) los diferentes contenidos disponibles en tiempos distintos para ser incorporados a la narración. ¿Ya estaba ahí esta sensación en el primer instante de la conciencia de la intención (en cuyo caso el efecto requiere un largo retraso para «mostrar el tiempo» y es estaliniano) o era una reinterpretación retrospectiva de lo que, sin esa reinterpretación, sería un *fait accompli* que se presta a confusión (y entonces sería orwelliano)? Espero que haya quedado claro que son las propias presuposiciones de la pregunta las que la descalifican.

6. Cabos sueltos

Es posible que usted aún quiera objetar que todos los argumentos presentados en este capítulo no son lo bastante poderosos como para derribar lo que no es más que la pura verdad, es decir, que nuestras experiencias de los acontecimientos se producen en el mismo orden en que las experimentamos. Si alguien piensa «uno, dos, tres, cuatro, cinco», el pensamiento «uno» se produce antes que el pensamiento «dos», y así sucesivamente. Este ejemplo ilustra una tesis que, por lo general, es correcta, y que parece no tener excepciones mientras limitemos nuestra atención a fenómenos psicológicos de dimensiones «ordinarias», es decir macroscópicas. Sin embargo, los experimentos que hemos examinado tienen relación con acontecimientos restringidos a marcos temporales excepcionalmente pequeños de unos pocos cientos de milisegundos. A esta escala, los supuestos tradicionales dejan de ser válidos.

Todo evento en su cerebro tiene una localización espacio-temporal definida, pero el preguntarse «en qué momento exactamente es *usted* consciente del estímulo» presupone que algunos de estos eventos son, o equivalen a, ser consciente del estímulo. Eso sería como preguntar: ¿en qué momento exactamente quedó informado el imperio británico de la tregua en la guerra de 1812? En algún momento entre el 24 de diciembre de 1814 y mediados de enero de 1815. Esto es bastante definido, pero no hay un hecho decisivo que nos permita precisar un día y una hora concretos. Aun en el caso de que

pudiéramos determinar los instantes precisos en que los distintos funcionarios y dirigentes del imperio fueron informados, ninguno de estos momentos puede ser considerado como el momento en que el imperio mismo fue informado. La firma de la tregua fue un acto oficial e intencionado del imperio, pero también lo fue la participación de las tropas británicas en la batalla de Nueva Orleans, y este último fue un acto llevado a cabo de acuerdo con el supuesto de que no se había firmado ninguna tregua. Se podría argumentar que el momento de la llegada de las noticias a Whitehall o al palacio de Buckingham en Londres debería ser considerado como el instante oficial en que el imperio quedó informado, ya que éste era el «centro nervioso» del imperio. Descartes pensó que la glándula pineal era precisamente el centro nervioso de nuestro cerebro, pero estaba equivocado. Dado que la cognición y el control —y, por tanto, la conciencia— se distribuyen por el cerebro, ningún momento puede ser considerado como el instante preciso en que se produce un evento consciente determinado.

En este capítulo he intentado hacer que algunos malos hábitos de pensamiento se tambaleen, a fin de privarles de sus imaginarios «cimientos» y de sustituirlos por mejores maneras de pensar, pero por el camino no he podido evitar algunos cabos sueltos. El más tentador de todos, creo, es la afirmación metafórica de que el «sondear» es un acto que «precipita narraciones». He afirmado que la organización temporal de los sondeos de los investigadores puede tener un efecto revisionista sobre los sistemas de representación utilizados por el cerebro. Pero entre los que pueden efectuar tales sondeos sobre el sujeto se encuentra el propio sujeto. Si usted se siente interesado por la cuestión de en qué momento se hace usted consciente de algo, sus propios autorreconocimientos o autoinvestigaciones fijan los límites de nuevas ventanas de control, alterando así las restricciones sobre los procesos que intervienen.

El resultado de los sondeos llevados a cabo por observadores exteriores son, normalmente, actos de habla de uno u otro tipo que *expresan juicios sobre* diferentes contenidos de la conciencia que cualquiera puede escuchar e interpretar. El resultado de los autosondeos son elementos *que pertenecen a la misma categoría semántica,* no son «presentaciones» (en el Teatro Cartesiano), sino *juicios* sobre lo que le parece al sujeto, juicios que el propio sujeto puede interpretar y recordar, y de acuerdo con ellos obrar. En ambos casos, estos eventos fijan interpretaciones de lo que experimentó el sujeto y, por tanto, establecen puntos fijos en la secuencia subjetiva. Sin embargo, en el modelo de las Versiones Múltiples no hay ninguna cuestión adicional sobre si *además* de dicho juicio, y de las discriminaciones previas sobre las que se basa, hubo una *presentación* de los materiales susceptibles de ser interpretados para su examen por parte de un juez supremo, la audiencia del Teatro Cartesiano. Ésta no es una idea fácil de comprender, y mucho menos de aceptar. Todavía nos quedan muchos caminos por construir que nos lleven hasta ella.

Capítulo 7

LA EVOLUCIÓN DE LA CONCIENCIA

Todo es lo que es porque se ha convertido en lo que es.
D'ARCY THOMPSON, *Sobre el crecimiento y la forma,* 1917

I. En la caja negra de la conciencia

La teoría esbozada en el capítulo anterior es un pequeño paso hacia la demostración de cómo puede residir la conciencia en el cerebro humano, aunque su principal contribución fue negativa: el derrocamiento de la idea del Teatro Cartesiano. Hemos empezado a sustituirla por una alternativa positiva, pero es cierto que no hemos ido muy lejos. A fin de hacer nuevos progresos, debemos cambiar de campo y aproximarnos a las complejidades de la conciencia desde una perspectiva diferente: la de la evolución. Ya que no siempre ha habido conciencia, ésta debe de haber surgido a partir de fenómenos previos que no podían ser considerados como conciencia. Quizá si examinamos lo que debe —o puede— haber participado en esta transición, tendremos una perspectiva mejor de tales complejidades y de su papel en el desarrollo de un fenómeno completo.

En su elegante librito, *Vehicles: Experiments in Synthetic Psychology* (1984), el investigador del cerebro Valentino Braitenberg describe una serie de mecanismos autónomos aún más complejos; partiendo de dispositivos de una simplicidad irrisoria y faltos de vida por completo ha ido construyendo gradualmente unas entidades (imaginarias) dotadas de una apariencia biológica y psicológica impresionante. Este ejercicio de imaginación funciona gracias a lo que él denomina la ley del análisis ascendente y la síntesis descendente: es más fácil imaginar la conducta (y las implicaciones comportamentales) de un dispositivo si lo sintetizamos «desde adentro hacia afuera», podríamos decir, que intentar analizar la conducta externa de una «caja negra» y adivinar lo que puede estar ocurriendo en su interior.

Hasta ahora hemos estando tratando la conciencia como si fuera algo parecido a una caja negra. Hemos considerado su «conducta» (= fenomenología) como algo «dado» y nos hemos preguntado qué tipo de mecanismos ocultos en el cerebro podían explicarla. Ahora vamos a invertir nuestra estrategia, y vamos a pensar en la evolución de los mecanismos cerebrales con ésta o aquella función, a ver si surge algo que nos proporcione un mecanis-

mo plausible para explicar algunas de las enigmáticas «conductas» que muestran nuestros cerebros conscientes.

Ha habido muchas teorías —de hecho, especulaciones— sobre la evolución de la conciencia humana, empezando por las conjeturas del propio Darwin en *El origen del hombre* (1871). A diferencia de la mayoría de explicaciones en ciencia, las explicaciones evolucionistas son, esencialmente, narraciones que nos llevan desde un tiempo en que algo no existía a un tiempo en que ese algo ya existía siguiendo una serie de pasos que la narración explica. En vez de adoptar una actitud escolar e intentar pasar revista a todas las narraciones que se han ideado, propongo contar una sola historia, tomando prestadas ideas de otros teóricos, y concentrándome en algunos puntos que a menudo se pasan por alto, y que nos ayudarán a superar los obstáculos que nos impiden comprender qué es la conciencia. En aras de contar una buena historia, y de que ésta sea relativamente breve, me he resistido a la tentación de incluir muchos y fascinantes episodios secundarios, y he reprimido el típico instinto de filósofo de airear todos los argumentos en favor y en contra de las ideas que incluyo y rechazo. El resultado, lo admito, es un poco como un resumen en cien palabras de *Guerra y paz*, pero tenemos mucho que hacer.[1]

La historia que contaremos se parece mucho a otras historias que la biología está empezando a contar. Comparémosla, por ejemplo, con la historia sobre los orígenes del sexo. Existen muchos organismos hoy en día que no tienen sexo y que se reproducen asexualmente, y hubo un tiempo en que todos los organismos existentes carecían de sexo, masculino y femenino. De alguna manera, a través de una serie de pasos cuya naturaleza nos podemos representar, algunos de estos organismos deben de haber evolucionado hasta tener sexo y, eventualmente, hasta convertirse en lo que nosotros somos ahora. ¿Qué condiciones se necesitaron para fomentar o hacer necesarias dichas innovaciones? ¿Por qué, en pocas palabras, se produjeron todos estos cambios? Éstos son algunos de los problemas más acuciantes de la teoría de la evolución contemporánea.[2]

Existe un hermoso paralelismo entre la pregunta que se plantea el origen del sexo y la que se plantea el origen de la conciencia. No hay nada de *erótico* (en términos humanos) en la vida sexual de las flores, las ostras u

1. Usted puede inferir que yo pienso que todo lo que utilizo en mi historia es correcto —o que va por el buen camino—, pero no puede inferir, del hecho de que no me ocupe de alguna teoría o de algún detalle de una teoría, que piense que es incorrecta. Tampoco puede usted inferir que yo piense que una teoría determinada es defendible, del hecho de que utilice algunos detalles de esta teoría. Lo dicho se aplica también a mis escritos anteriores sobre este asunto, de los cuales extraigo algunas ideas.

2. John Maynard Smith es el principal teórico que se ha ocupado de estos asuntos y, además de su clásico *The Evolution of Sex* (1978), el lector hallará muchos artículos brillantes que tratan los problemas conceptuales de la teoría en la recopilación *Sex, Games, and Evolution* (1989). Véase también R. Dawkins (1976), págs. 46-48 [págs. 55-57 de la edición castellana; N. del T.], para un breve repaso de estos temas.

otras formas simples de vida, pero podemos reconocer en sus mecánicas y aparentemente gozosas rutinas de reproducción los fundamentos y los principios de nuestro, mucho más apasionante, mundo del sexo. De forma similar, no hay nada particularmente *egoico* (si se me permite acuñar el término) en los primitivos precursores de los yos humanos conscientes, aunque sientan las bases a partir de las que se han desarrollado las innovaciones y complejidades humanas.* El diseño de nuestras mentes conscientes es el resultado de tres procesos evolutivos sucesivos, apilados el uno sobre el otro, y, a fin de poder comprender esta pirámide de procesos, tenemos que empezar por el principio.

2. El principio

Escena primera: el nacimiento de los límites y de las razones

Al principio no había razones; sólo había causas. Nada tenía un propósito, ni nada tenía algo que pudiera denominarse una función; en el mundo no había teleología. La explicación de todo esto es simple: no había nada que tuviera intereses. Pero después de varios milenios aparecieron los *replicadores* simples (R. Dawkins, 1976; véase también Monod, 1972, cap. 1, «Extraños objetos»). Aunque ellos no tuvieran el menor indicio de cuáles eran sus intereses y, quizá, no se pueda realmente decir que tuvieran intereses, nosotros, oteando hacia atrás, a esos días remotos, desde nuestra atalaya de pequeños dioses, podemos, sin peligro de caer en la arbitrariedad, atribuirles ciertos intereses, producto de su definido «interés» en la autorreplicación. Es decir, quizá no había diferencia, no había ningún objetivo implicado, a nadie ni a nada importaba si conseguían replicarse o no (aunque parece que podemos estar agradecidos de que así fuera), pero al menos podemos, condicionalmente, asignarles ciertos intereses. *Si* estos replicadores simples tenían que sobrevivir y replicarse, persistiendo ante una entropía creciente, su entorno debía cumplir ciertas condiciones: las condiciones propicias para la replicación debían estar presentes o ser, cuando menos, frecuentes.

Por expresarlo de manera más antropomórfica, si los replicadores simples quieren seguir replicándose, deben esperar y afanarse por conseguir varias cosas; deben evitar las cosas «malas» y buscar las cosas «buenas». Cuando aparece en escena una entidad capaz de mostrar la conducta, por muy primitiva que ésta sea, de evitar su disolución y su descomposición, ésta trae al mundo todo lo que tiene de «bueno». Es decir, crea un punto de vista desde el cual los acontecimientos del mundo pueden clasificarse

* En la traducción de los términos «sexy» y «selfy» que utiliza Dennett, me permito la libertad de tomar prestados del traductor francés de esta obra Pascal Engel (que, a su vez, los toma del italiano Lauro Colasanti) los términos «erótico» y «egoico».

en favorables, desfavorables y neutrales. Y son sus propias tendencias innatas a buscar los primeros, alejarse de los segundos e ignorar los terceros las que contribuyen de forma determinante a la definición de estas tres clases. A medida que la criatura empieza a tener intereses, el mundo y sus acontecimientos empiezan a crear *razones* para ella, independientemente de si la criatura puede reconocerlas en su totalidad o no (Dennett, 1984a). Las primeras razones preexistieron a su reconocimiento. Evidentemente, el primer problema al que se enfrentaron los primeros afrontadores de problemas fue el de aprender a reconocer y actuar de acuerdo con las razones que su propia existencia había hecho nacer.

Tan pronto como algo se pone manos a la obra en la tarea de la autoconservación, los límites empiezan a cobrar importancia, ya que si usted se ha propuesto luchar por su propia conservación, no quiere malgastar esfuerzos intentando conservar el resto del mundo: traza usted una línea. En una palabra, usted se hace *egoísta*. Esta forma primordial de egoísmo (que, en tanto que forma primordial, carece de los muchos matices que poseen *nuestras* formas de egoísmo) es uno de los signos distintivos de la vida. El punto en que un pedazo de granito acaba y en que empieza el siguiente pedazo es insignificante; la frontera marcada por la fractura puede ser real, pero no hay nada que trabaje por la protección del territorio, por empujar la frontera o por preparar la retirada. «Yo contra el mundo» —esa distinción entre todo lo que está dentro de unos límites y todo lo que está en el mundo exterior— se halla en el centro de todo proceso biológico, y no solamente en la ingestión y la excreción, la respiración y la transpiración. Considérese, por ejemplo, el sistema inmunológico, con sus millones de anticuerpos distintos prestos a la defensa del cuerpo contra millones de intrusos. Este ejército debe resolver el problema del reconocimiento: distinguirse a sí mismo (y a sus compañeros) de todo lo demás. El problema se ha resuelto de una manera bastante parecida a como los Estados y sus ejércitos han resuelto su problema homólogo: mediante rutinas de identificación estandarizadas y mecanizadas; los pasaportes y los funcionarios de aduanas en miniatura son formas moleculares y detectores de formas. Es importante reconocer que este ejército de anticuerpos no tiene generales, ni cuartel general con un plan de batalla, ni siquiera una descripción del enemigo: los anticuerpos representan a sus enemigos de la misma manera que un millón de cerrojos representan a las llaves que los abren.

Es preciso mencionar otros hechos que ya son evidentes en este estadio inicial. Aunque la evolución dependa de la historia, la Madre Naturaleza no es una esnob y los orígenes no tienen nada que hacer con ella. No importa dónde ni cómo un organismo haya adquirido su destreza; tanto haces, tanto vales. Por lo que sabemos, claro está, los pedigríes de los primeros replicadores eran todos más o menos los mismos: cada uno de ellos era el producto de una u otra serie ciega y aleatoria de selecciones. Pero si un hiperingeniero, viajando en el tiempo, hubiera insertado un *replicador robot* en ese medio, y si sus destrezas hubieran sido iguales o mejores que las de sus compe-

tidores surgidos naturalmente, sus descendientes podrían estar entre nosotros, ¡incluso podríamos *ser* nosotros mismos! (Dennett, 1987a, 1990b).

La selección natural no puede saber cómo un sistema llegó a ser lo que es, lo cual no significa que no pueda haber profundas diferencias entre los sistemas «diseñados» por selección natural y los diseñados por ingenieros inteligentes (Langton, Hogeweg, en Langton, 1989). Por ejemplo, los diseñadores humanos, que son previsores pero un poco cabeza cuadrada, ven que sus diseños a veces se echan perder a causa de efectos secundarios e interacciones imprevistos, así que intentan guardarlos de tales incidencias concediendo a cada elemento del sistema una única función, y aislándolo de los demás elementos. Por el contrario, la Madre Naturaleza (el proceso de selección natural) es famosa por su miopía y su falta de objetivos. Como no prevé nada, no tiene manera de preocuparse por los efectos secundarios imprevistos. Como no «intenta» evitarlos, prueba diseños que tienen muchos efectos secundarios; muchos de estos diseños son terribles (pregúntenle a un ingeniero), pero de vez en cuando se produce un *efecto secundario inesperado*: dos o más sistemas funcionales sin relación interactúan para producir una bonificación: funciones múltiples para elementos únicos. Las funciones múltiples no son desconocidas en los mecanismos producto de la ingeniería humana, pero son bastante raras; en la naturaleza están por todas partes y, como veremos, uno de los motivos por los cuales los teóricos han tenido tantos problemas para hallar diseños aceptables de la conciencia en el cerebro es precisamente que han tendido a pensar en los elementos cerebrales como poseedores de una única función cada uno.[3]

Hemos puesto las primeras piedras. Ahora podemos explicar los hechos primordiales siguientes:

3. La idea de las neuronas multifuncionales no es nueva, pero sólo recientemente ha empezado a ganar adeptos:

> Son las concatenaciones más o menos simultáneas de salidas o señales neuronales las que son ambiguas, y no las salidas de neuronas individuales. La convergencia de diversas concatenaciones de señales ambiguas en cada nivel sucesivo resolvería la ambigüedad de la misma manera que la convergencia de definiciones ambiguas determina una única solución (o casi única) en un crucigrama (Dennett, 1969, pág. 56).

> ...no hay una estructura o combinación de grupos única correspondiente a una categoría o patrón de salida dado. Por el contrario, más de una combinación de grupos neuronales puede dar lugar a una salida determinada, y un único grupo determinado puede participar en más de un tipo de función señaladora. Esta propiedad de los grupos neuronales en repertorio, denominada *degenerancia*, constituye uno de los fundamentos de las capacidades generalizadoras de los mapas de reentrada (Edelman, 1989, pág. 50).

Sobre este rasgo arquitectónico, en el que cada nodo contribuye a muchos contenidos diferentes, ya había insistido Hebb en su trabajo pionero *The Organization of Behavior: A Neuropsychological Theory* (1949). Es la base del «procesamiento paralelo distribuido» o «conexionismo». Pero hay algo más sobre la multiplicidad de funciones: en un nivel de análisis más burdo, veremos sistemas enteros con funciones especializadas, pero también capaces de participar en proyectos más generales.

1. Hay razones que reconocer.
2. Allí donde hay razones, hay puntos de vista desde los cuales reconocerlas y evaluarlas.
3. Todo agente debe distinguir el «aquí dentro» del «mundo exterior».
4. Todo acto de reconocimiento debe, en última instancia, ser puesto en práctica por una miríada de rutinas «ciegas y mecánicas».
5. Dentro de los límites defendidos, no tiene por qué haber un alto mando o un cuartel general.
6. En la naturaleza, tanto haces, tanto vales; los orígenes no cuentan.
7. En la naturaleza, los elementos cumplen a menudo funciones múltiples dentro de la economía de un único organismo.

Ya hemos escuchado los ecos de estos hechos primordiales, en la búsqueda del último «punto de vista del observador consciente» y en todos aquellos casos en que hemos sustituido homúnculos por (equipos de) mecanismos simples. Pero, como hemos visto, el punto de vista del observador consciente no es idéntico a los puntos de vista primordiales de los primeros replicadores que dividieron su mundo en lo bueno y lo malo, sino que es un descendiente evolucionado de esos puntos de vista. (Después de todo, incluso las plantas tienen puntos de vista en este sentido primordial.)

Escena segunda: nuevas y mejores maneras de producir el futuro

> Y una de las funciones más profundas y más generales de los organismos vivientes es la de mirar hacia adelante, la de producir futuro, como dijo Paul Valéry.
>
> FRANÇOIS JACOB, *The possible and the actual*, 1982, pág. 66

> Predecir el futuro de una curva comporta llevar a cabo una determinada operación en el pasado. La verdadera operación predictiva no puede ser llevada a cabo por ningún aparato que se pueda construir, pero existen ciertas operaciones que se le parecen y que, de hecho, pueden ser ejecutadas por aparatos que podemos construir.
>
> NORBERT WIENER, *Cibernética*, 1948, pág. 12

En el capítulo anterior hice mención, aunque de soslayo, del hecho de que el propósito fundamental de los cerebros es el de producir futuro. Esta afirmación merece que le dediquemos un poco más de atención. Para poder arreglárselas, un organismo o bien debe acorazarse (como los árboles o las almejas) y «esperar que no pase lo peor», o bien debe desarrollar métodos de alejarse del peligro hacia áreas más seguras del vecindario. Si usted elige el segundo camino, entonces deberá enfrentarse con el problema primordial que todo *agente* tiene que estar resolviendo continuamente:

¿Y ahora qué hago?

A fin de resolver este problema, usted necesita un sistema nervioso para *controlar* sus actividades en el tiempo y en el espacio. La ascidia, cuando es joven, vaga por el mar en busca de una buena roca o arrecife de coral al que agarrarse para convertirlo en su casa por el resto de su vida. Para llevar a cabo esta tarea, posee un sistema nervioso rudimentario. Cuando encuentra el sitio y hecha raíces, ya no necesita más a su cerebro, así que ¡se lo come! (Es como ganar unas oposiciones a cátedra.)[4] La clave del control es la capacidad de *rastrear* e incluso *anticipar* los rasgos más importantes del entorno, de modo que todos los cerebros son, en esencia, *máquinas de anticipación*. La concha de la almeja es un buen blindaje, pero no siempre puede estar cerrada; el reflejo preconfigurado que cierra rápidamente la concha es un burdo pero efectivo mecanismo para anticipar/evitar peligros.

Aún más primitivas son las respuestas de retirada y acercamiento de los organismos más simples, las cuales están ligadas de manera todavía más directa con las fuentes de lo que es bueno y lo que es malo: las *tocan*. Así, dependiendo de si la cosa tocada es buena o mala para ellos, estos organismos retroceden o engullen (en el último momento, si tienen suerte). Y actúan de este modo simplemente porque su «configuración» está instalada de tal manera que es el contacto real con un rasgo bueno o malo lo que desencadena el movimiento reflejo. Como veremos, este hecho es la base de las características más terribles y más deliciosas (literalmente) de la conciencia. Al principio, *todas* las «señales» causadas por entidades en el entorno significaban o bien «¡lárgate!» o bien «¡ve a por ello!» (Humphrey, 1992).

Ningún sistema nervioso, en esos tiempos tan tempranos, tenía la posibilidad de utilizar un «mensaje» más desapasionado u objetivo que se limitara a *informarlo*, de manera neutral, de alguna condición. Sin embargo, dichos sistemas nerviosos no pueden sacar mucho partido del mundo. Sólo son capaces de lo que podríamos denominar *anticipación proximal*: aquella conducta que es apropiada para lo que se va a producir en un futuro *inmediato*. Los mejores cerebros son aquellos que pueden extraer más información, más rápidamente, y utilizarla para *evitar* todo contacto nocivo desde el principio, o para *buscar* alimento (y las oportunidades de aparearse, una vez el sexo ha aparecido).

Enfrentados a la tarea de extraer futuro útil a partir de nuestros propios pasados, nosotros los organismos intentamos conseguir algo gratis (o, cuando menos, rebajado): descubrir las leyes del mundo —y si no existen, descubrir algo que se aproxime a las leyes del mundo—, cualquier cosa que nos proporcione un límite. Desde ciertas perspectivas parece realmente sorprendente que nosotros los organismos obtengamos algo de la naturaleza. ¿Hay algún motivo por el cual la naturaleza deba mostrarse tal como es, revelar sus regularidades cuando alguien la observa? Todo productor de futuro útil suele ser algo así como una estratagema, un sistema improvisado que resul-

4. Esta analogía entre la ascidia y el catedrático fue observada por primera vez, creo, por el investigador del cerebro Rodolfo Llinás.

ta funcionar con más frecuencia de lo que sería de esperar, una diana en una regularidad del mundo que se puede rastrear. Todos estos afortunados anticipadores con los que se tropieza la Madre Naturaleza están, evidentemente, destinados a recibir su recompensa, si consiguen mejorar los límites del organismo.

En el extremo más bajo de la escala, pues, tenemos criaturas que representan lo mínimo posible: lo suficiente como para *permitir al mundo advertirles* cuando empiezan a hacer algo malo. Las criaturas que adoptan esta política no suelen seguir ningún plan. Se tiran de cabeza, y si algo duele «saben lo suficiente» como para retirarse, pero eso es lo mejor que pueden hacer.

El siguiente paso comporta la anticipación de corto alcance como, por ejemplo, la capacidad de esquivar obstáculos. Este talento anticipatorio a menudo está «preconfigurado»; forma parte de la maquinaria innata diseñada a lo largo de eones para detectar el tipo de regularidad (con excepciones) que podemos observar entre *cosas que aparecen repentinamente y cosas que nos golpean*. La respuesta de esquivar amenazas está preconfigurada en los seres humanos, por ejemplo, y puede observarse en los recién nacidos (Yonas, 1981), un regalo de nuestros remotos antepasados cuyos primos, que no sobrevivieron, no sabían esquivar lo bastante bien. ¿Podemos decir que la señal «¡una amenaza!» *significa* «¡esquiva!»? Digamos que lo protosignifica; está directamente instalado en el mecanismo de esquivar.

Tenemos otros regalos parecidos. Nuestros sistemas visuales, como los de muchos otros animales, los peces incluidos, son extremadamente sensibles a patrones con un eje de simetría vertical. Braitenberg sugiere que ello se debe probablemente al hecho de que en el mundo natural de nuestros remotos antepasados (mucho antes de que hubiera fachadas de iglesias y puentes colgantes), las únicas cosas que había en el mundo que *mostraran* ejes de simetría verticales eran otros animales, y sólo cuando estaban de frente, de cara a uno. Así que nuestros antepasados se equiparon con uno de los más valiosos sistemas de alarma, que se disparaba (principalmente) cada vez que otro animal los *miraba* (Braitenberg, 1984).[5] Identificar a un depredador a una cierta distancia (en el espacio), en vez de tener que esperar a sentir sus dientes clavándose en el cuerpo, es también un modo de anticipación distal *en el tiempo*: concede una cabeza de ventaja para evitar el obstáculo.

Un aspecto importante de estos mecanismos es lo burdo de su capacidad de discriminación; cambian lo que podríamos denominar *veracidad y precisión en el testimonio* por *la velocidad y la economía*. Algunas de las cosas que disparan el detector de simetrías verticales no tienen una particular significación para el organismo: los raros casos de árboles o arbustos

5. Esta configuración recuerda vagamente a la manera, no totalmente infalible, que tenía Shakey de distinguir cajas de pirámides. Así que Shakey no es totalmente antibiológico después de todo; la biosfera posee muchos artilugios como éstos. También es verdad, sin embargo, que el sistema «visual» de Shakey no es un modelo particularmente bueno de la visión de ninguna especie viviente. No era éste su objetivo.

casi simétricos o (en tiempos más modernos) muchos artefactos humanos. Así pues, la clase de cosas que el mecanismo distingue es, oficialmente, una pandilla multicolor, dominada por animales mirando-en-mi-dirección, pero dejando abierta la posibilidad de disparar un buen número de falsas alarmas (relativas a *ese* mensaje). Y ni siquiera todos o sólo los patrones con simetría vertical pondrán el dispositivo en funcionamiento; algunos patrones verticalmente simétricos no conseguirán, por un motivo u otro, disparar la alarma e, incluso en este caso, también habrá falsas alarmas; es el precio que hay que pagar por un mecanismo rápido, barato y transportable, un precio que los organismos, con su narcisismo (Akins, 1989), pagan a gusto. Todo esto es fácil de ver, pero algunas de sus implicaciones con respecto a la conciencia no son tan obvias a primera vista. (En el capítulo 12 cobrará importancia cuando planteemos preguntas tales como ¿qué propiedades detectamos con nuestra visión en color?, ¿qué tienen en común las cosas rojas? E, incluso, ¿por qué el mundo se nos aparece como se nos aparece?)

Llegar a estar informado (faliblemente) de que un animal le está mirando es casi siempre un acontecimiento con cierta significación en el mundo natural. Si el animal no lo quiere devorar, puede que sea un compañero potencial, o un rival para competir por un compañero, o una presa que ha detectado su presencia. Acto seguido, la alarma debe poner en funcionamiento los analizadores «¿amigo, enemigo o comida?» a fin de que el organismo pueda distinguir entre mensajes tales como: «¡un miembro de tu especie te está mirando!», «¡un depredador te tiene en su línea de mira!» y «¡tu cena está a punto de escaparse!». En algunas especies (ciertos peces, por ejemplo) el detector de simetrías verticales se ha configurado de tal manera que provoca una inmediata interrupción de toda actividad en curso conocida como *respuesta de orientación*.

El psicólogo Odmar Neumann (1990) sugiere que las respuestas de orientación son el equivalente biológico de la alarma de barco «¡Todo el mundo al puente!». La mayoría de animales, como nosotros, desarrolla actividades que éstos controlan de forma rutinaria, «con piloto automático», haciendo uso de mucho menos del total de sus capacidades bajo el control, de hecho, de subsistemas especializados en sus cerebros. Cuando se dispara una alarma especializada (como nuestra alarma para amenazas o la de simetrías verticales), o se dispara una alarma general ante la presencia de algo repentino y sorprendente (o solamente inesperado), el sistema nervioso del animal se moviliza a fin de enfrentarse a la posibilidad de una emergencia. El animal deja de hacer lo que estaba haciendo y lleva a cabo una rápida exploración o puesta al día que concede a cada uno de los órganos de los sentidos la oportunidad de contribuir al conjunto de información disponible y relevante. Se establece un centro de control *temporal* a través de un aumento de la actividad neuronal; por un breve espacio de tiempo, todas las líneas permanecen abiertas. Si el resultado de esta puesta al día es que se dispara una «segunda alarma», entonces se moviliza todo el cuerpo del animal con una descarga de adrenalina. Si no es así, la actividad disminuye, la tripulación sin obli-

gaciones se va a dormir, y los especialistas reanudan sus funciones de control. Estos breves episodios de interrupción e intensa vigilancia no son en sí mismos episodios que se correspondan al estilo humano de «atención consciente» (como apuntan insistentemente algunas personas) o, cuando menos, no son necesariamente ejemplos de este estado, pero probablemente son los necesarios precursores, en la evolución, de nuestros estados conscientes.

Neumann especula sobre la idea de que estas respuestas de orientación se iniciaron como reacciones a señales de alarma, pero que se mostraron tan útiles al provocar una puesta al día generalizada, que los animales empezaron a entrar en el modo de orientación cada vez con más frecuencia. Sus sistemas nerviosos necesitaban un a modo de «¡todos al puente!», pero, una vez lo obtuvieron, el coste de ponerlo en funcionamiento con mayor frecuencia era muy pequeño, si no nulo, y en cambio resultaba beneficioso al mejorar la información sobre el estado del entorno o sobre el estado del propio animal. Podríamos decir que se convirtió en un hábito, ya no bajo el control de estímulos externos, sino desencadenado internamente (un poco como los ensayos de alarma de incendio).

De la vigilancia regular se pasó gradualmente a la *exploración* regular y así empezó a evolucionar una nueva estrategia comportamental: la estrategia de adquirir información «por sí misma», por si podía ser útil en algún momento. La mayoría de los mamíferos fueron conducidos hasta esta estrategia, especialmente los primates, que desarrollaron ojos con una gran movilidad, los cuales, con sus movimientos sacádicos, llevaban a cabo una exploración casi ininterrumpida del mundo. Esto marcó un cambio fundamental en la economía de los organismos que dieron este salto: el nacimiento de la curiosidad o apetito epistémico. En vez de recoger información sólo según el principio del pague-ahora-y-consuma-ahora, empezaron a convertirse en lo que el psicólogo George Miller ha dado en llamar *informívoros*: organismos hambrientos de información sobre el mundo que habitan (y sobre sí mismos). Sin embargo, ni inventaron ni desplegaron sistemas totalmente nuevos de recogida de la información. Como suele ocurrir en la evolución, ensamblaron estos nuevos mecanismos a partir de los mecanismos que les había proporcionado su herencia. Esta historia ha dejado sus huellas, particularmente en los matices emocionales o afectivos de la conciencia, pues, aunque las criaturas superiores se habían convertido en recolectores «desinteresados» de información, sus «reporteros» no eran más que los viejos chivatos y animadores de sus antepasados, que nunca enviaban un mensaje «directo», sino añadiendo siempre un «giro» positivo o negativo a toda la información que proporcionaban. Eliminando las comillas y las metáforas: los lazos innatos entre estados informativos y actos de retirada o engullimiento, evitación y refuerzo no se rompieron, sólo se atenuaron y se redirigieron. (Volveremos sobre este punto en el capítulo 12.)

En los mamíferos, este desarrollo evolutivo se vio potenciado por la división del trabajo en el cerebro, que dio lugar a dos áreas especializadas: (aproximadamente) la *dorsal* y la *ventral*. (Lo que sigue es una hipótesis del neu-

ropsicólogo Marcel Kinsbourne.) El cerebro dorsal recibió responsabilidades de *pilotaje* «directo» a fin de mantener la nave (el cuerpo del organismo) alejada de los peligros; al igual que los controles de «detección de colisiones» de los videojuegos, tenía que estar casi continuamente explorando, en busca de cosas que se acercaban o alejaban, y, en general, era el responsable de evitar que el organismo chocara con cosas o se despeñara por un barranco. Esto dejaba al cerebro ventral con un poco de tiempo libre para concentrarse en la identificación de los distintos objetos del mundo; podía focalizar sobre casos particulares y analizarlos de manera serial y con una relativa lentitud, ya que podía confiar en el sistema dorsal para mantener la nave alejada de los escollos. De acuerdo con las especulaciones de Kinsbourne, en los primates dicha especialización dorsal-ventral se vio alterada hasta dar lugar a la tan celebrada especialización hemisferio-derecho/hemisferio-izquierdo: el hemisferio derecho, global y espaciotemporal, y el hemisferio izquierdo, más concentrado, analítico y serial.

Sólo hemos examinado una de las ramas en la historia evolutiva de los sistemas nerviosos, y hemos utilizado uno de los más básicos mecanismos evolutivos: la selección de *genotipos* (combinaciones de genes) determinados que probadamente han dado lugar a individuos mejor adaptados (*fenotipos*) que los genotipos alternativos. Los organismos que tienen la suerte de tener una buena configuración en el momento de su nacimiento tienden a tener proles con más altos índices de supervivencia, de modo que las buenas configuraciones se extienden entre las poblaciones. También hemos esbozado una progresión en cuanto al diseño, desde los más simples detectores imaginables de lo bueno y lo malo hasta conjuntos de tales mecanismos organizados en una arquitectura con una capacidad considerable de producir anticipación útil en entornos relativamente estables y predictibles.

Para la siguiente fase de nuestra historia, debemos introducir una innovación mayor: la emergencia de fenotipos individuales cuyo interior no se halla completamente configurado, sino que es variable o *plástico*, y que, por lo tanto, pueden aprender a lo largo de su vida. La emergencia de la plasticidad en los sistemas nerviosos se produjo al mismo tiempo (aproximadamente) que los desarrollos que acabamos de esbozar, y estableció dos nuevos medios en los cuales se podía producir la evolución con mucha más rapidez que la evolución genética que procede, sin ayuda alguna, a través de la mutación de los genes y la selección natural. Habida cuenta de que algunas de las complejidades de la conciencia humana son el resultado de desarrollos que se produjeron, y que continúan produciéndose, en estos nuevos medios, necesitamos una caracterización clara, aunque elemental, de las relaciones entre el uno y el otro y con el proceso subyacente de la evolución genética.

3. Evolución en los cerebros y el efecto Baldwin

Todos asumimos que el futuro será igual que el pasado. Como señaló Hume, ésta es la premisa esencial, aunque imposible de probar, de todas nues-

tras inferencias inductivas. La Madre Naturaleza (ese diseñador-constructor reflejado en los procesos de selección natural) adopta el mismo supuesto. Desde muchos puntos de vista, las cosas siempre son las mismas: la gravedad sigue ejerciendo su fuerza, el agua sigue evaporándose, los organismos siguen necesitando reponer y proteger el agua de sus cuerpos, las cosas amenazadoras siguen subtendiendo porciones cada vez mayores de las retinas, etc. Allí donde todas estas generalidades son importantes, la Madre Naturaleza proporciona soluciones a largo plazo para los problemas: detectores preconfigurados basados en la gravedad para saber dónde es arriba y dónde es abajo, alarmas también preconfiguradas para la sed, y circuitos preconfigurados para esquivar-cuando-algo-nos-amenaza. Hay otras cosas que cambian, pero de forma previsible, en ciclos, y la Madre Naturaleza responde a ellas con otros mecanismos preconfigurados, tales como dispositivos para el desarrollo de pelajes de invierno desencadenados por los cambios de temperatura, y relojes interiores para regular el despertar y los ciclos de sueño de los animales nocturnos y diurnos. Pero, a veces, las oportunidades y las vicisitudes del entorno son relativamente imprevisibles, para la Madre Naturaleza y para cualquier otro, pues son procesos *caóticos* o están influidos por procesos de este tipo (Dennett, 1984a, págs. 109 y sigs.). En estos casos, ningún diseño estereotipado podrá adaptarse a todas las eventualidades, de modo que los mejores organismos serán aquellos capaces de *rediseñarse a sí mismos* en mayor o menor grado a fin de enfrentarse a las condiciones que encuentren. En ciertas ocasiones este rediseño se denomina *aprendizaje*, en otras se denomina simplemente desarrollo. La frontera es difícil de fijar. ¿*Aprenden* a volar los pájaros? ¿Aprenden a cantar sus canciones? (Nottebohm, 1984; Marler y Sherman, 1983) ¿*Aprenden* a hacer crecer sus plumas? ¿*Aprenden* a andar o a hablar los bebés? Puesto que la frontera (si es que la hay) no es particularmente relevante para nuestros objetivos, nos referiremos a cualquiera de estos procesos, desde aprender-a-enfocar-los-ojos a aprender-mecánica-cuántica, con el término de *fijación postnatal del diseño*. Cuando nacemos, queda un cierto espacio para la variación, que finalmente se fija mediante un proceso u otro en un elemento de diseño relativamente permanente para el resto de su vida (una vez hemos aprendido a montar en bicicleta o a hablar ruso, ese conocimiento tiende a quedarse con nosotros).

¿De qué manera puede producirse un proceso de fijación postnatal del diseño? Sólo de una manera (no milagrosa): mediante un proceso muy parecido al proceso que fija el diseño prenatal o, en otras palabras, mediante un proceso de evolución por selección natural que se lleva a cabo en el individuo (en el fenotipo). Algo previamente fijado en el individuo a través de la selección natural es lo que debe jugar el papel de selector mecánico, y otras cosas deben jugar el papel de candidatos a la selección. Se han propuesto muchas teorías diferentes de este proceso, pero todas ellas —con la excepción de las que son extravagantes o particularmente misteriosas— poseen esta estructura, y difieren únicamente en los detalles relacionados con los mecanismos propuestos. Durante gran parte del siglo XX, la teoría más in-

fluyente ha sido el conductismo de B. F. Skinner, de acuerdo con la cual los emparejamientos de estímulos con respuestas jugaban el papel de candidatos para la selección, mientras que los estímulos «de refuerzo» eran el mecanismo de selección. No se puede negar que los estímulos placenteros y los estímulos dolorosos —el palo y la zanahoria— juegan un papel en la formación de la conducta, pero es también un hecho ampliamente aceptado que el «condicionamiento operante» del conductismo es un mecanismo demasiado simple para explicar las complejidades de la fijación postnatal del diseño en especies tan complejas como la humana (y, probablemente, también en las palomas, pero éste es otro problema). Hoy se suelen favorecer aquellas teorías que sitúan el proceso evolutivo en el cerebro (Dennett, 1974). Desde hace décadas han circulado diferentes versiones de esta idea y, ahora, con la posibilidad de evaluar modelos rivales mediante simulaciones por ordenador, el debate se ha enardecido, por lo que procuraremos mantenernos al margen del mismo.[6]

Para nuestros propósitos basta con decir que, de una manera u otra, el cerebro plástico es capaz de reorganizarse a sí mismo adaptativamente en respuesta a determinadas novedades que encuentra el organismo en su entorno, y que el proceso mediante el cual el cerebro hace esto es, con toda certeza, un proceso mecánico muy parecido a la selección natural. Éste es el primer nuevo medio de evolución: la fijación postnatal del diseño en cerebros individuales. Los candidatos para la selección son ciertas estructuras cerebrales que controlan o influyen en la conducta, y la selección se lleva a cabo a través de un proceso mecánico de eliminación que, a su vez, está genéticamente instalado en el sistema nervioso.

Sorprendentemente, esta capacidad, también producto de la evolución por selección natural, no sólo da una ventaja a los organismos que la poseen sobre sus primos preconfigurados que no se pueden diseñar a sí mismos, sino que también realimenta el proceso de evolución genética y *lo acelera*. Éste es un fenómeno que ha recibido varias denominaciones, aunque la más conocida es la de efecto Baldwin (véase Richards, 1987; Schull, 1990). Así es como funciona.

6. Las ideas básicas ya pueden hallarse en los escritos del propio Darwin y de sus primeros discípulos (Richards, 1987). El neuroanatomista J. Z. Young (1965a, 1965b) fue el primero en proponer una teoría seleccionista de la memoria (véase también Young, 1979). Yo mismo me he ocupado de desarrollar una versión filosófica del mismo argumento, con un pequeño esbozo de los detalles relevantes, en mi tesis doctoral defendida en Oxford en 1965, una versión mejorada de la cual es «Evolution in the Brain», el capítulo 3 de *Content and Consciousness*, 1969. John Holland (1975) y otros dedicados a la inteligencia artificial han desarrollado «algoritmos genéticos» para sistemas de aprendizaje o de autorrediseño (véanse también Holland, Holyoak, Nisbett y Thagard, 1986), y Jean-Pierre Changeux (Changeux y Danchin, 1976; Changeux y Dehaene, 1989), desarrollaron un modelo neuronal bastante detallado. El neurobiólogo William Calvin (1987, 1989a) aporta una perspectiva diferente (y más fácilmente accesible) sobre el asunto en sus propias teorías de la evolución en el cerebro. Véase también su clara y penetrante reseña (Calvin, 1989b) del libro *Neural Darwinism* de Gerald Edelman (1987). Recientemente, Edelman ha publicado *The Remembered Present: A Biological Theory of Consciousness* (1989).

Considérese una población de una determinada especie en la que se observa una variación considerable en cuanto a cómo están configurados los cerebros de los nuevos miembros que nacen. Supongamos que una de esas posibles configuraciones dota al que la posee con un buen truco, un talento comportamental que lo protege o que hace aumentar sus oportunidades de forma espectacular. Esto podemos representarlo en lo que se denomina paisaje adaptativo; la altura representa idoneidad (a mayor altura, mayor grado de idoneidad) y la longitud y la latitud representan variables en la configuración (no es preciso que las especifiquemos para este experimento mental).

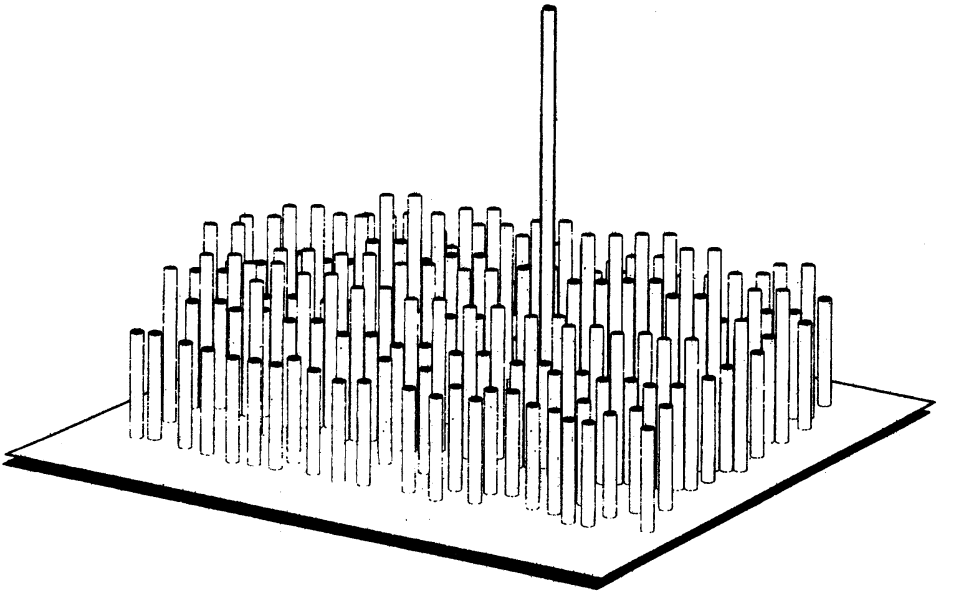


Figura 7.1

Como muestra la figura, sólo una de las configuraciones es la favorecida; las otras, no importa lo «cerca» que estén de la configuración buena, tienen un grado de idoneidad casi idéntico. Una aguja en el pajar como ésta puede ser prácticamente invisible para la selección natural. Incluso si algunos individuos afortunados poseen esta configuración, las probabilidades de que su suerte se extienda por la población de generaciones siguientes pueden ser ínfimas *a menos que* haya plasticidad de diseño en los individuos.

Supóngase, pues, que todos los individuos empiezan siendo diferentes genéticamente, pero en el curso de su vida, gracias a su plasticidad, deambulan por el espacio de posibilidades de diseño que les son accesibles. Y, dadas las circunstancias particulares del entorno, todos tienden a girar alrededor de la configuración favorecida. Hay un buen truco que aprender en su entorno, y todos tienden a aprenderlo. Supóngase que es un truco tan bueno que los que nunca lo aprenden se hallan en clara desventaja, y supóngase también que aquellos que nunca lo aprenden son los que empiezan su vida con diseños que están muy alejados, en el espacio de diseños posibles, de los que están cerca del buen truco (y, por tanto, necesitan mucho más diseño postnatal).

Una fantasía (adaptada de Hinton y Nowlan, 1987) nos ayudará a imaginarlo. Supóngase que hay diez puntos en el cerebro de cada animal donde se puede conectar un «cable» de dos maneras distintas, A o B. Supóngase que el buen truco es el diseño cuya configuración es AAABBBAAAA, y que las demás configuraciones son, desde el punto de vista de la conducta, igual de poco interesantes. Dado que todas estas conexiones son plásticas, cada animal, en el curso de su vida, puede intentar cualquiera de las 2^{10} combinaciones distintas en la configuración de A y B. Aquellos animales que han nacido en estados tales como BAABBBAAAA están a un paso del buen truco (aunque es evidente que pueden perderse en una serie de intentos erróneos). Otros, cuya configuración inicial es BBBAAABBBB necesitan un mínimo de diez pasos (en el supuesto de que nunca se equivoquen al reconfigurar) antes de dar con el buen truco. Aquellos animales cuyos cerebros empiezan más cerca del objetivo tendrán una ventaja en cuanto a su supervivencia frente a los que empezaron más lejos, incluso en el caso de que no exista ninguna *otra* ventaja selectiva por haber nacido con una configuración «que falla por poco» frente a haber nacido con una configuración «que falla por mucho» (como se demuestra en la figura 7.1). Así, la población de la siguiente generación tenderá a contener un mayor número de individuos con configuraciones próximas al objetivo (y, por tanto, más capaces de alcanzarlo en el tiempo de su vida), y el proceso seguirá hasta que toda la población haya fijado *genéticamente* el buen truco. Por lo tanto, un buen truco «descubierto» así por los individuos de una población puede pasar con relativa rapidez a las generaciones futuras.

Si otorgamos a los individuos una posibilidad variable de acertar el buen truco (y, por tanto, «de reconocerlo» y «de aferrarse a él») en el curso de sus vidas, la casi invisible aguja en un pajar de la figura 7.1 se convierte en la cumbre de una colina que la selección natural puede escalar (figura 7.2). Este proceso, el efecto Baldwin, podría en un principio parecernos la tan denostada idea lamarckiana de la transmisión genética de las características adquiridas, pero no lo es. Nada de lo que el individuo aprende es transmitido a su prole. Se trata simplemente del hecho de que los individuos que son lo bastante afortunados como para nacer más cerca en el espacio de exploración de diseño a un buen truco aprendible, tenderán a tener mayor des-

cendencia, y ésta, a su vez, tenderá a estar más cerca del buen truco. Con el paso de las generaciones, la competencia se hace más dura: llega un momento en que, *a menos que* uno no haya nacido con el buen truco (o muy cerca de él), no estará lo bastante cerca para competir. Si no fuera por la plasticidad, no obstante, el efecto no se produciría, ya que «un fallo por poco o un fallo por mucho siempre es un fallo» *a menos que* uno pueda ir intentando nuevas variaciones hasta dar con la configuración acertada.

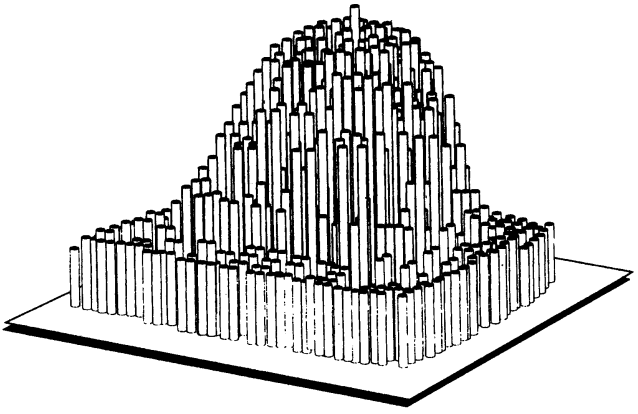


Figura 7.2

Gracias al efecto Baldwin, podemos decir que las especies evalúan con antelación la eficacia de diseños diferentes concretos a través de la exploración fenotípica (individual) del espacio de las posibilidades cercanas. Si se descubre una configuración vencedora determinada, tal descubrimiento *creará* una nueva presión selectiva: los organismos que, en el paisaje adaptativo, estén más cerca de este descubrimiento tendrán una clara ventaja sobre aquellos que estén más lejos. Ello significa que las especies con plasticidad *tenderán* a evolucionar más deprisa (y con mayor «lucidez») que las que no la tienen. Así pues, la evolución en el primer medio, la plasticidad fenotípica, puede potenciar la evolución en el segundo medio, la variación genética. (Enseguida consideraremos un efecto compensatorio que surge como resultado de las interacciones con el tercer medio.)

4. Plasticidad en el cerebro humano: montando el decorado

> [A]sí también el entendimiento, con su fuerza natural, se forja instrumentos intelectuales, con los que adquiere nuevas fuerzas para realizar otras obras intelectuales y con éstas consigue nuevos ins-

trumentos, es decir, el poder llevar más lejos la investigación, y sigue así progresivamente, hasta conseguir la cumbre de la sabiduría.
BARUCH SPINOZA, *Tratado de la reforma del entendimiento*, 1677*

Los sistemas nerviosos preconfigurados son ligeros, eficientes desde el punto de vista energético y útiles para organismos que deben arreglárselas en entornos estereotípicos y que disponen de un presupuesto limitado. Los cerebros más imaginativos, gracias a su plasticidad, son capaces no sólo de mostrar anticipación estereotípica, sino también de ajustarse a las tendencias. Incluso el modesto sapo posee un pequeño grado de libertad para responder a la novedad, alterando lentamente sus pautas de acción a fin de seguir la pista —con un retraso considerable— a los cambios de aquellas características del entorno que pueden afectar a su bienestar (Ewert, 1987). En el cerebro del sapo, un diseño para tratar con el mundo evoluciona a un ritmo mucho más rápido que la selección natural, en el que las «generaciones» duran segundos o minutos, no años. Sin embargo, para tener un control realmente potente, lo que se necesita es una máquina de anticipación capaz de realizar ajustes mayores en unos pocos milisegundos, y para eso se necesita un virtuoso de la producción de futuro, un sistema con capacidad de previsión, que evite la rutina en su propia actividad, que resuelva problemas antes de encontrárselos, y que reconozca nuevos indicadores de lo bueno y de lo malo. Con toda nuestra estupidez, nosotros los seres humanos estamos mucho mejor equipados para estas tareas que cualquier otro organismo autocontrolador, y es nuestro enorme cerebro el que lo hace posible. Pero, ¿cómo?

Hagamos un repaso de nuestros progresos. Hemos esbozado la historia —sólo un hilo de esa tela multidimensional que es la historia evolutiva— de la evolución del cerebro de los primates. Al estar basado en milenios de sistemas nerviosos anteriores, consta de un conglomerado de circuitos especializados diseñados para llevar a cabo tareas determinadas en la economía de sus otros antecesores primates: detectores de cosas amenazadoras conectados a mecanismos para esquivarlas, detectores alguien-me-está-mirando conectados a discriminadores amigo-o-enemigo-o-comida, a su vez conectados con las subrutinas apropiadas. A todo ello, podemos añadir aquellos circuitos específicos de los primates tales como los circuitos para la coordinación de ojos y manos diseñados para recolectar bayas y recoger semillas, así como otros para agarrarse a las ramas de los árboles, e incluso otros para tratar con objetos situados a poca distancia de la cara (Rizzolati, Gentilucci y Matelli, 1985). Gracias a unos ojos móviles y a una inclinación por la exploración y la puesta al día, estos cerebros de primate se veían inundados con regularidad por información multimedia (o, como diría un investigador del cerebro, información multimodal), con lo que se les planteó un nuevo problema: el del control a alto nivel.

* Trad. cast. de A. Domínguez, Madrid, Alianza, 1988 (§31, págs. 86-87). [N. del T.]

Un problema es también una oportunidad, una puerta abierta hacia un nuevo espacio de diseño. Podemos suponer que hasta ahora los sistemas nerviosos resolvían el problema del «¿y ahora qué hago?» buscando el equilibrio relativamente simple entre un repertorio estrictamente limitado de acciones, compuesto, si no por las que responden al acrónimo CLEF (comer, luchar, escapar y aparearse),* cuando menos por una versión más o menos modesta de las mismas. Pero ahora, con el aumento de la plasticidad funcional, y de la cantidad de información «centralizada» disponible para los distintos especialistas, el problema de qué hacer después engendró un metaproblema: el de *en qué pensar* después. Está muy bien equiparse con una subrutina de «¡todos al puente!», pero entonces, una vez todos están en el puente, es preciso disponer de un medio con el que enfrentarse a la avalancha de voluntarios. No es de suponer que ya hubiera un capitán dispuesto a hacerlo (¿qué habría estado haciendo hasta entonces?), así que los conflictos entre voluntarios deben de haberse resuelto sin la mediación de un alto mando. (Como vimos en el ejemplo del sistema inmunológico, la acción organizada y coordinada no tiene por qué depender siempre del control desde un mando central.) El modelo original para este tipo de procesos es el de la arquitectura en pandemónium de la inteligencia artificial propuesto por Oliver Selfridge (1959), en el que muchos «demonios» compiten en paralelo por la hegemonía. Dado que el nombre con que Selfridge bautizó a este tipo de arquitecturas me parece muy apropiado, en este libro lo utilizaré con un sentido genérico para referirme al propio modelo de Selfridge así como a todos sus descendientes, directos e indirectos, como por ejemplo la «programación de conflictos» (Norman y Shallice, 1980; Shallice, 1988) o las redes del quien-gana-se-queda-con-todo, de Ballard y Feldman (1982), y los sistemas derivados de ellas.

La programación de conflictos siguiendo el modelo del pandemónium, al estar dirigida de forma bastante directa por las características actuales del entorno, sigue dando lugar a sistemas nerviosos con una capacidad de previsión limitada. Del mismo modo que Odmar Neumann hipotetizó que las reacciones de orientación, originalmente guiadas por las novedades del entorno, se iniciaron de manera endógena (desde dentro), nosotros podemos lanzar la hipótesis de que se produjo una presión para desarrollar una manera más endógena de resolver el metaproblema de en qué pensar después, una presión para crear algo en el interior con algo más que los poderes organizativos imaginarios de un capitán.

Considérese cómo podría ser la conducta de ese hipotético primate antecesor nuestro vista desde fuera (por el momento posponemos para más tarde toda cuestión relacionada con qué se sentiría en caso de ser ese primate): un animal capaz de aprender trucos nuevos, casi constantemente vigilante y sensible a la novedad, pero con una «capacidad de atención escasa» y una

* Aquí procuro ser fiel al estilo desenfadado del autor intentando adaptar al castellano la pequeña frivolidad que éste se permite en el texto original. [N. del T.]

tendencia de dicha capacidad de atención a ser «captada» por cualquier característica del entorno que pudiera distraerlo. Nada de proyectos a largo plazo para este animal, como mínimo nada de proyectos nuevos. (Debemos dejar espacio para subrutinas estereotípicas de larga duración configuradas genéticamente como, por ejemplo, la rutina para la construcción de nidos de los pájaros, la de construcción de presas de los castores, o la de encontrar alimento de los pájaros y las ardillas.)

Sobre el sustrato de este sistema nervioso, ahora queremos imaginarnos la construcción de una mente más humana, con algo parecido a un «flujo de la conciencia» capaz de sostener el tipo de «cadenas de pensamientos» de las que aparentemente depende la civilización humana. Los chimpancés son nuestros parientes más cercanos —genéticamente, están más cerca de nosotros que de los gorilas o de los orangutanes— y hoy en día existe el convencimiento de que tenemos un antepasado común que vivió hace unos seis millones de años. Desde esa primera ruptura, nuestros cerebros se han diferenciado de manera espectacular, aunque más por su tamaño que por su estructura. Mientras que los chimpancés tienen un cerebro con un tamaño muy parecido al del de nuestro antepasado común (y es importante —y difícil— no olvidar que los chimpancés también han evolucionado a partir de ese antepasado común), los cerebros de nuestros antepasados homínidos se hicieron cuatro veces más grandes. Este aumento de volumen no se produjo inmediatamente; durante muchos millones de años después de la separación de los protochimpancés, nuestros antepasados homínidos siguieron adelante con cerebros de simio en cuanto al tamaño se refiere, a pesar de haberse convertido en bípedos hace ya unos tres millones y medio de años. Entonces, cuando empezaron las glaciaciones, hace unos dos millones y medio de años, comenzó la gran encefalización, que no se completó hasta hace unos 150.000 años, *antes* del desarrollo del lenguaje, el cocinar los alimentos, y la agricultura. Por qué los cerebros de nuestros antepasados crecieron tanto y tan deprisa (en la escala evolutiva es más una explosión que un florecimiento) sigue siendo motivo de debate y controversias (para algunos análisis bastante clarificadores, véanse los libros de William Calvin). Pero existe total acuerdo en cuanto a la naturaleza del producto: el cerebro del primer *Homo sapiens* (que vivió desde hace unos 150.000 años hasta el final de la última glaciación hace sólo 10.000 años) era un cerebro terriblemente complejo de una plasticidad sin rival, apenas distinguible del nuestro en cuanto a tamaño y forma. Esto es importante: el sorprendente crecimiento del cerebro del homínido, se había completado esencialmente *antes* del desarrollo del lenguaje, de modo que el hecho de que el lenguaje se hiciera posible no puede ser la respuesta a las complejidades de la mente. Las especializaciones innatas para el lenguaje, cuya existencia, propuesta por el lingüista Noam Chomsky y otros, empieza a confirmarse en ciertos detalles neuroanatómicos, es un añadido muy reciente y apresurado, y, sin duda, fruto de la explotación de circuitos secuenciales anteriores (Calvin, 1989a), acelerada por el efecto Baldwin. Además, la expansión más notable de los poderes *mentales*

humanos (como lo atestigua el desarrollo de la cocina, la agricultura, el arte y, en una palabra, la civilización) se ha producido aún más recientemente, desde el final de la última glaciación, en un suspiro de 10.000 años que, desde la perspectiva evolutiva que mide las tendencias en una escala de millones de años, es tanto como decir que ha sido casi instantánea. Al nacer, nuestros cerebros están equipados con sólo unas pocas facultades más de las que tenía el cerebro de nuestros antepasados hace 10.000 años. Así pues, el impresionante avance del *Homo sapiens* en los últimos 10.000 años tiene que ser debido casi por entero al aprovechamiento de la plasticidad de ese cerebro de formas radicalmente originales: creando algo parecido a un *software* para potenciar sus facultades subyacentes (Dennett, 1986).

En breve, nuestros antepasados deben de haber aprendido algunos buenos trucos que pudieran llevar a cabo con su *hardware* ajustable, y que nuestra especie, gracias al efecto Baldwin, está empezando a transferir al genoma. Además, como veremos, existen motivos para creer que a pesar de la presión selectiva inicial en favor de una «preconfiguración» gradual de estos buenos trucos, éstos han alterado tanto la naturaleza del entorno para nuestra especie que apenas queda ya presión selectiva en favor de ulteriores preconfiguraciones. Es posible que casi toda la presión selectiva sobre el diseño del desarrollo del sistema nervioso humano se haya detenido a causa de los efectos secundarios provocados por esa nueva oportunidad de diseño que explotaron nuestros antepasados.

Hasta ahora me he cuidado mucho de hablar de estos sistemas nerviosos simples como sistemas capaces de *representar* algo en el mundo. Los diferentes diseños que hemos considerado, tanto plásticos como preconfigurados, pueden ser vistos como sensibles a, interesados en, diseñados-con-un-ojo-en, o capaces de utilizar la información sobre las diferentes características del entorno del organismo y, por tanto, en este sentido más restringido podrían llamarse representaciones; pero ahora debemos detenernos un momento a considerar qué rasgos de estos diseños complejos podrían llevarnos a considerarlos como sistemas de representaciones.

Parte de la variabilidad en el cerebro es necesaria simplemente como medio para los patrones transitorios de actividad cerebral que *registran* o, en cualquier caso, *siguen la pista* a las características variables relevantes del entorno. Algo en el cerebro tiene que variar a fin de registrar que un pájaro ha pasado volando, o que la temperatura del aire ha bajado, o que alguno de los estados del organismo se ha visto alterado (por ejemplo, un descenso de la concentración de azúcar en la sangre, un aumento de dióxido de carbono en los pulmones). Además —y éste es el punto de apoyo que le da su fuerza a las genuinas representaciones—, estos patrones transitorios internos llegan a ser capaces de continuar «siguiendo la pista» (en un sentido amplio del término) a los rasgos a que se refieren cuando se interrumpe la interacción causal con sus referentes. «Una cebra que ha avistado un león no olvida dónde está el león cuando deja de mirar hacia él por un momento. El león tampoco olvida dónde está la cebra» (Margolis, 1987, pág. 53). Compárese

lo dicho con el fenómeno más simple del girasol que sigue el paso del sol por el cielo, ajustando su ángulo como un panel solar orientable a fin de maximizar la cantidad de luz solar que recibe. Si el sol se ve oscurecido temporalmente, el girasol no es capaz de proyectar su trayectoria; el mecanismo que es sensible al paso del sol no representa el paso del sol en el sentido amplio que apuntábamos. Los principios de la verdadera representación se encuentran en numerosos animales inferiores (y no deberíamos eliminar, *a priori*, la posibilidad de que haya representaciones reales entre las plantas), pero en los seres humanos dicha capacidad de representar se ha disparado hacia arriba.

Entre las cosas que un humano adulto en cierta medida puede representar se hallan, no sólo:

1. la posición de su cuerpo y sus extremidades;
2. un punto de luz roja;
3. un cierto grado de hambre;
4. un cierto grado de sed;
5. el aroma de un buen Borgoña añejo;

sino también:

6. el aroma de un buen Borgoña añejo como el aroma de un Chambertin de 1971;
7. París;
8. la Atlántida;
9. la raíz cuadrada del mayor número primo menor que 20;
10. el concepto de la combinación de un sacacorchos niquelado y un quitagrapas.

Es casi seguro que ningún cerebro de cualquier animal es capaz de representar los puntos que van del 6 al 10, y también es seguro que se necesita un largo proceso de ajuste antes de que el cerebro de un niño sea capaz de registrar o representar cualquiera de estas cosas. Por el contrario, las primeras cinco pueden perfectamente ser cosas que casi cualquier cerebro podría representar (en algún sentido) sin necesidad de entrenamiento previo.

En cualquier caso, la manera en que el cerebro representa el hambre debe de diferir, físicamente, de cómo representa la sed, ya que debe regular conductas distintas dependiendo de qué se representa. En el otro extremo de la cadena, también debe de haber una diferencia entre la manera que un ser humano adulto representa París o la Atlántida, ya que pensar en la primera no es igual que pensar en la segunda. ¿Cómo es posible que un determinado estado o evento cerebral represente una característica del mundo y no otra?[7] Y sea lo que sea lo que hace que un cierto rasgo cerebral represente lo que re-

7. Éste es el problema, fundamental en la filosofía de la mente, del *contenido mental* o de la *intencionalidad*, para el cual se han propuesto numerosas soluciones, todas ellas controvertidas. La mía se desarrolla en *La actitud intencional* (1987a).

presenta, *¿cómo llega* a representar lo que representa? De nuevo (¡espero que esta muletilla no empiece a resultar pesada!), existe un amplio abanico de posibilidades establecidas por los procesos evolutivos: algunos de los elementos del sistema de representación pueden estar —de hecho, deben ser (Dennett, 1969)— determinados de forma innata, y el resto deben ser «aprendidos». Mientras que algunas de las categorías relevantes de la vida (como el hambre y la sed) sin duda nos vienen «dadas» por el modo en que estamos configurados al nacer, otras tenemos que desarrollarlas por nosotros mismos.[8]

¿Cómo lo hacemos? Probablemente a través de un proceso de generación y selección de patrones de actividad neuronal en el córtex cerebral, ese enorme manto lleno de circunvoluciones que ha crecido con gran rapidez dentro del cráneo humano y que actualmente cubre por completo al viejo cerebro animal. Con sólo decir que se trata de un proceso evolutivo que ha tenido lugar principalmente en el córtex dejamos de lado muchas cosas que siguen siendo un misterio, y a este nivel de complejidad y sofisticación, incluso si conseguimos explicar el proceso en el nivel de las sinapsis o de los grupos neuronales, seguiremos desorientados sobre otros aspectos de lo que tiene que estar ocurriendo. Si queremos encontrarle un sentido a todo esto, primero debemos situarnos en un nivel más general y abstracto. Una vez tengamos una mayor comprensión, por vaga que sea, de los procesos en este nivel más alto, podremos pensar en volver a bajar al nivel más mecánico del cerebro.

La plasticidad es lo que hace posible el aprendizaje, pero será mejor que en nuestro entorno haya *algo que aprender* que a su vez sea el producto de un proceso previo de diseño, a fin de que cada uno de nosotros no se dedique a reinventar la rueda cada vez. La evolución cultural, y la transmisión de sus productos, es el segundo medio de evolución, y ésta depende de la plasticidad fenotípica tanto como la plasticidad fenotípica depende de la variación genética. Nosotros los seres humanos hemos utilizado nuestra plasticidad no sólo para aprender, sino también para aprender a cómo aprender mejor, y después hemos aprendido mejor a cómo aprender mejor a cómo

8. Algunos osados teóricos afirman lo contrario. Jerry Fodor (1975), por ejemplo, sostiene que todos los conceptos que uno puede llegar a poseer son innatos, aunque sólo se activan o podemos acceder a ellos a través de determinadas experiencias «de aprendizaje». Así, Aristóteles poseía en su cerebro el concepto de avión, y también el de bicicleta, pero nunca tuvo ocasión de utilizarlos. A todos aquellos que se echan a reír ante semejante idea tan descabellada, Fodor replica que los inmunólogos solían reírse de la idea de que las personas —Aristóteles, por ejemplo— nacen con millones de anticuerpos distintos, incluidos los anticuerpos para determinados compuestos que han surgido en la naturaleza sólo en el siglo XX, pero ya no se ríen. El problema con esta idea, en su aplicación tanto a la inmunología como a la psicología, es que en sus versiones más radicales es claramente falsa y que las versiones más moderadas son difíciles de distinguir de la tesis opuesta. En el sistema inmunológico, *hay* una reacción combinatoria —no *toda* respuesta inmunológica entre tipos de anticuerpos preexistentes es biunívoca; igualmente, quizás Aristóteles poseyera un concepto innato de *avión*, pero ¿poseía también el concepto innato de *jumbo-jet*? ¿Y qué podemos decir del concepto de *billete de ida y vuelta con tarifa APEX de Boston a Londres*? Una vez se empieza a encontrar una respuesta para este tipo de preguntas, en ambos campos, resulta que parece que es preciso apelar al aprendizaje en mayor o en menor grado, y apelar al innatismo de ciertos conceptos.

aprender mejor, y así sucesivamente. También hemos aprendido a hacer que el fruto de este aprendizaje sea accesible a los recién llegados. En cierto modo, *instalamos* un *sistema* de hábitos previamente inventado y «depurado» en un cerebro parcialmente estructurado.

5. La invención de buenos y malos hábitos de autoestimulación

> ¿Cómo puedo decir lo que pienso antes de ver lo que digo?
> E. M. FORSTER, 1960

> Hablamos no sólo para contar lo que pensamos a los demás, sino también para contarnos lo que pensamos a nosotros mismos.
> J. HUGHLINGS JACKSON, 1915

¿Cómo puede haber llegado a producirse esta posibilidad de compartir el *software*? Una primera respuesta que diga «porque sí» abre el camino para una respuesta más elaborada. Considérese el momento en la historia del primer *Homo sapiens* en que el lenguaje —o quizá debamos llamarlo protolenguaje— estaba empezando a desarrollarse. Estos antepasados nuestros eran bípedos omnívoros que vivían en pequeños grupos de individuos emparentados entre sí, y que probablemente habían desarrollado hábitos de vocalización con propósitos determinados muy parecidos a los que observamos hoy en día en los chimpancés y los gorilas, así como en otras especies más alejadas como el *Cercopithecus pygerythrus** (Cheney y Seyfarth, 1990). Podemos suponer que los actos comunicativos (o cuasicomunicativos) ejecutados por estas vocalizaciones no merecían todavía el calificativo de actos de habla (Bennett, 1976), en los que la intención del emisor de producir un cierto efecto en la audiencia depende de la capacidad de esa audiencia de apreciar dicha intención.[9] Podemos suponer, no obstante, que estos antepasados, al igual que los primates actuales capaces de emitir vocalizaciones, eran capaces de distinguir emisores y audiencias diferentes en situaciones diferentes, utilizando información acerca de lo que ambas partes pudieran creer o querer.[10] Por ejemplo, el homínido Alf no se tomaría la molestia de hacer

* Con el nombre genérico de mono verde se designa en castellano una especie de monos del género *Cercopithecus* muy común en África Central. Los zoólogos distinguen tres subespecies de mono verde: el *Cercopithecus pygerythrus* («vervet monkey» en inglés), el *Cercopithecus aethiops* («grivet monkey» en inglés) y el *Cercopithecus sabaeus* («green monkey» en inglés). [N. del T.]

9. Evidentemente, estoy haciendo alusión a la teoría del significado no natural de Paul Grice (Grice, 1957, 1969), pero para una teoría de la comunicación que elimina algunos de los aspectos más débiles y menos verosímiles de las teorías griceanas, véase Sperber y Wilson (1986).

10. ¿Qué derecho tengo a hablar de creencias y deseos al referirme a estos antepasados nuestros todavía no completamente conscientes? Mi propia teoría de las creencias y los deseos, expuesta en *La actitud intencional*, defiende la visión de que no hay ninguna razón de peso para escribir estos términos entre comillas: la conducta de los animales «inferiores» (incluso las ranas) es un dominio de explicación tan bueno para la actitud intencional, con su atribución de creencias y deseos, como lo es la conducta de los seres humanos. En cualquier caso, los lectores que no estén de acuerdo con esta teoría pueden interpretar el uso que aquí se hace de estos términos como un uso metafórico y ampliado de los mismos.

que el homínido Bob creyera que no había comida en la cueva (gruñendo algo así como «nohaypapeo»), si Alf creyera que Bob ya sabía que no había comida en la cueva. Y si Bob pensaba que Alf quería engañarle, Bob haría bien de tomarse la vocalización de Alf con cierto escepticismo.[11]

Siguiendo con nuestras especulaciones, a veces, al hallarse uno de estos homínidos encallado en algún proyecto, se vería en la necesidad de «pedir ayuda» y, en particular, de «pedir información». En unas ocasiones, la audiencia presente respondería «comunicando» algo que tuviera los efectos deseados sobre el solicitante, haciéndole probar una vía alternativa o haciéndole «ver» una solución para su problema. A fin de que esta práctica acabara por arraigar en una comunidad, los solicitantes deberían ser capaces, en justa reciprocidad, de asumir el papel de contestadores. Deberían poseer una capacidad comportamental que se pondría en marcha ocasionalmente emitiendo proferencias «útiles» como reacción a proferencias con la función de «peticiones» emitidas por los demás. Por ejemplo, si un homínido sabía algo y se le «preguntaba» sobre ello, normalmente esto podía tener el efecto de hacerle «decir lo que sabía».

En otras palabras, estoy proponiendo que hubo un tiempo en la evolución del lenguaje en que las vocalizaciones tenían la función de solicitar y compartir información útil, aunque no tenemos por qué asumir que un espíritu cooperativo de ayuda mutua tuviera un valor de supervivencia, ni que habría sido un sistema estable si hubiera surgido. (Véase, Dawkins, 1982, págs. 55 y sigs.; véase también Sperber y Wilson, 1986.) Por el contrario, debemos asumir que los costes y los beneficios fruto de participar de dicha práctica eran en cierta medida «visibles» para estas criaturas, y que un número suficiente de ellas vio que los beneficios *para sí mismas* superaban a los costes, de modo que los hábitos de comunicación acabaron por quedar establecidos dentro de la comunidad.

Entonces un buen día (en esta reconstrucción racional), uno de estos homínidos «por error» pidió ayuda en un momento en que nadie que pudiera ayudarle se encontraba presente con la excepción de... ¡él mismo! Al oír su propia petición, la estimulación provocó el mismo tipo de proferencia para ayudar a los demás que hubiera causado en nuestro homínido una petición por parte de otro homínido. Así que, para regocijo de nuestra criatura, ésta descubrió que había conseguido que ella misma respondiera a su propia pregunta.

Lo que estoy intentando justificar con este experimento mental, en el que deliberadamente he simplificado muchos detalles, es la afirmación de que la práctica de hacerse preguntas a uno mismo pudo surgir de forma natural como efecto secundario de la práctica de hacer preguntas a los demás, y que su utilidad sería parecida: sería un conducta que podría ser reconocida como

11. Sobre la comunicación en los primates y sobre ciertos aspectos todavía no resueltos acerca de la presunta capacidad de engañar deliberadamente en los simios y los monos, véanse Dennett (1983, 1988c, 1988d, 1989a); Byrne y Whiten (1988); Whiten y Byrne (1988).

potenciadora de la expectativas de un individuo al promover un control más eficaz de las acciones. Todo lo que se necesita para que esta práctica tenga su utilidad es que las relaciones de acceso preexistentes en el cerebro de un individuo sean algo menos que óptimas. Supóngase, en otras palabras, que aunque la información adecuada para un fin determinado ya está *en el cerebro*, está en manos del especialista equivocado; el subsistema en el cerebro que necesita la información no puede obtenerla directamente del especialista, porque la evolución simplemente no ha llegado a establecer dicha «conexión». Sin embargo, el acto de provocar al especialista para «difundir» la información por el entorno y después confiar en la existencia de un par de orejas (y un sistema auditivo) para captarla, sería un modo de construir una «conexión virtual» entre ambos subsistemas.[12]

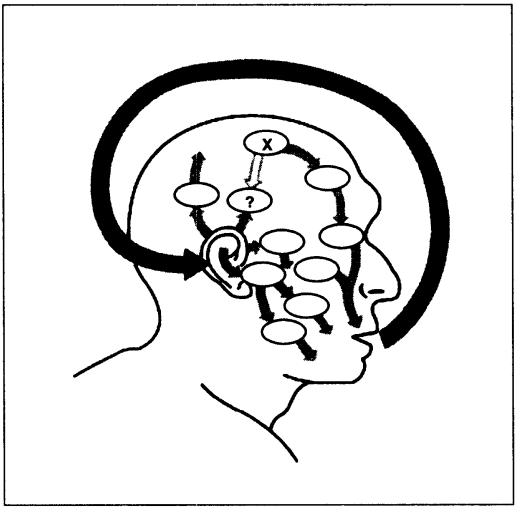


Figura 7.3

Dicho acto de autoestimulación podría abrir una nueva vía entre los componentes internos de un individuo. Dicho crudamente, el canalizar información por los oídos y el sistema auditivo de un individuo puede muy bien producir la estimulación del tipo de conexiones que éste buscaba, puede desencadenar los mecanismos asociativos adecuados, y puede, en fin, arrancarle el pedacito mental correcto de la punta de la lengua. Uno puede decirlo, oírse decirlo, y así obtener la respuesta que buscaba.

12. En «El jardín de senderos que se bifurcan», Jorge Luis Borges (1941) idea una diabólica e inteligente versión de esta estrategia, que me voy a abstener de describir por no privar al lector de un gran final.

Una vez estos toscos hábitos de autoestimulación oral empezaron a establecerse como buenos trucos en la conducta de las poblaciones de homínidos, sería de esperar que pronto empezaran a refinarse, tanto en los hábitos comportamentales aprendidos de la población y, gracias al efecto Baldwin, en las predisposiciones genéticas y en mejoras posteriores de eficiencia y efectividad. En particular, podemos especular que se reconocerían las virtudes de hablar *sotto voce* con uno mismo, conduciendo finalmente a un discurso enteramente silencioso dirigido a uno mismo. Este proceso silencioso mantendría el bucle de autoestimulación, pero desecharía las partes periféricas de vocalización y audición del proceso, cuya contribución no era tan importante. Tal innovación traería el beneficio adicional, oportunamente sancionado, de obtener una cierta privacidad durante la práctica de la autoestimulación cognitiva. (En el próximo capítulo consideraremos cómo podrían funcionar estas líneas de comunicación acortadas.) Dicha privacidad sería particularmente útil cuando hubiera miembros de la misma especie que pudieran escuchar. Esta conducta del hablar privadamente con uno mismo podría no ser la mejor manera imaginable de modificar la arquitectura funcional del cerebro de uno, pero sería una mejora accesible y fácil de descubrir, lo cual es más que suficiente. Sería lenta y laboriosa, comparada con los rápidos procesos cognitivos en los que estaba basada, porque debería utilizar largas porciones de sistema nervioso «diseñadas para otros propósitos», particularmente para la producción y la comprensión del habla. Sería igualmente lineal (limitada a un asunto por vez), como la comunicación social a partir de la cual evolucionó. Y dependería, al menos al principio, de las categorías informativas encarnadas por las acciones que explotaba. (Si sólo había cincuenta cosas que un homínido podía «decir» a otro homínido, entonces sólo había cincuenta cosas que se podía decir a sí mismo.)

Hablar en voz alta es sólo una posibilidad. Hacerse dibujos es otro acto de automanipulación fácil de apreciar. Supóngase que un día uno de estos homínidos casualmente dibujara dos líneas paralelas en el suelo de su cueva, y al ver lo que había hecho, esas dos líneas le recordaron, visualmente, las márgenes paralelas del río que tendría que cruzar más tarde ese día, y esto a su vez le recordó que tenía que llevar consigo una cuerda para cruzarlo. Si no hubiera hecho el «dibujo», podemos suponer que habría caminado hasta el río y entonces *se habría* dado cuenta, después de echar un vistazo rápido, de que necesitaba su cuerda y de que tenía que volver atrás para buscarla. Ello significaría un notable ahorro de tiempo y energía que desarrollaría nuevos hábitos y acabaría por refinarse bajo la forma de un hacer dibujos *privados* «en el ojo de la mente».

El talento humano para inventar nuevas vías de comunicación interna si la ocasión lo exige se revela de manera particularmente clara en casos de lesión cerebral. La gente tiene unas capacidades extraordinarias para superar las lesiones cerebrales, y nunca se trata de «sanar» ni de reparar los circuitos dañados. Al contrario, descubren nuevas maneras de llevar a cabo los viejos trucos, y la exploración activa juega un papel fundamental en la

rehabilitación. Encontramos una anécdota particularmente sugestiva en el estudio de los pacientes con el cerebro dividido (Gazzaniga, 1978). Los hemisferios derecho e izquierdo normalmente están conectados por una amplia franja de fibras denominada cuerpo calloso. Cuando ésta franja se secciona quirúrgicamente (en el tratamiento de algunas epilepsias graves), ambos hemisferios pierden la mayoría de sus «cables» de interconexión y quedan prácticamente incomunicados. Si a uno de estos pacientes se le pide que identifique un objeto —un lápiz, por ejemplo—, buscando en una bolsa y tanteándolo, el éxito depende de qué mano se utiliza par hacer la exploración táctil. La mayor parte del cableado en el cuerpo está conectada *contralateralmente*, con el hemisferio izquierdo recibiendo información de la parte derecha del cuerpo (y controlándola), y viceversa. Dado que es el hemisferio izquierdo el que normalmente controla el lenguaje, cuando el paciente utiliza la mano derecha para rebuscar en la bolsa no tiene ninguna dificultad en decir qué hay en su interior, pero si utiliza la mano izquierda, entonces el hemisferio derecho recibe la información de que el objeto es un lápiz, pero no tiene el poder de dirigir la voz para expresarlo. Ocasionalmente, sin embargo, parece que un hemisferio derecho da con una estrategia inteligente: al encontrar la punta del lápiz y clavársela en la palma de la mano, causa una señal de dolor agudo que viaja hacia arriba por el brazo izquierdo, y algunas fibras del dolor están conectadas *ipsilateralmente*. Así, el hemisferio izquierdo, que controla el lenguaje, tiene una pista: es algo lo bastante afilado como para causar dolor. «Es afilado, ¿será un bolígrafo? ¿Un lápiz, quizás?» El hemisferio derecho, al escuchar esta proferencia, puede ayudar aportando algunas pistas de más —fruncir el entrecejo para *bolígrafo*, sonreír para *lápiz*—, de modo que después de un breve turno de «veinte preguntas» el hemisferio izquierdo consigue dar con la respuesta correcta. Son muchas las anécdotas sobre los ingeniosos apaños que inventan para la ocasión los pacientes con el cerebro dividido, pero debemos tomarlas con la debida cautela. Podrían ser lo que parecen ser: casos en los que se muestra la habilidad con que el cerebro puede descubrir y poner en práctica estrategias de autoestimulación a fin de mejorar sus comunicaciones internas en ausencia de las conexiones «deseadas». Pero también podrían ser fantasías, fabricadas inconscientemente, de unos investigadores que esperan precisamente encontrar este tipo de evidencias. Éste es el problema que tienen las anécdotas.

Podríamos seguir divirtiéndonos imaginando otras situaciones posibles para explicar la «invención» de modos útiles de autoestimulación, pero ello podría ocultar el hecho de que no todas estas invenciones tendrían por qué ser útiles para sobrevivir. Una vez que el hábito general de la autoestimulación exploratoria se hubo inculcado de una u otra manera, éste podría haber engendrado una serie de variaciones no funcionales (aunque no necesariamente disfuncionales). Existen, después de todo, muchas variedades de autoestimulación y automanipulación que seguramente no tienen ningún efecto útil sobre la cognición o sobre el control, pero que, por razones dar-

winianas estándar, no pueden ser eliminadas e incluso pueden derivar hacia la fijación definitiva (cultural o genética) en ciertas subpoblaciones. Algunos candidatos posibles serían pintarse de azul, flagelarse con ramas de abedul, escarificarse, ayunar, repetir una fórmula «mágica» una y otra vez, o mirarse el ombligo. Si estas prácticas son hábitos que merece la pena inculcar, sus virtudes en tanto que potenciadores de la idoneidad no son lo bastante «obvias» como para elevarlas al rango de ninguna de las predisposiciones genéticas conocidas, pero quizá son invenciones demasiado recientes.

Las variedades de la autoestimulación que ahora potencian la organización cognitiva son probablemente en parte innatas y en parte aprendidas e idiosincrásicas. Del mismo modo que uno puede descubrir que al acariciarse de una determinada manera puede producir ciertos efectos secundarios deseables que son sólo parcial e indirectamente controlables —y entonces puede dedicar cierto tiempo e ingenio a desarrollar y explorar las técnicas para producir esos efectos secundarios—, uno puede explorar semiconscientemente ciertas técnicas de autoestimulación cognitiva, desarrollando un estilo personal con sus pros y sus contras. Algunas personas son mejores que otras en este tipo de actividades, y algunas nunca llegan a aprender los trucos, aunque hay mucho que compartir y que enseñar. La *transmisión* cultural, al permitir el acceso de todo el mundo a un buen truco, puede allanar la cumbre de la colina de idoneidad (véase la figura 7.2, pág. 200), creando una meseta que amaina la presión selectiva por transferir el truco al genoma. Si casi todo el mundo se hace lo bastante bueno como para arreglárselas en el mundo civilizado, la presión selectiva por transferir buenos trucos al genoma desaparece o se ve sustancialmente reducida.

6. El tercer proceso evolutivo: memas y evolución cultural[13]

Del mismo modo que aprendimos a ordeñar las vacas, y después a domesticarlas en nuestro beneficio, también hemos aprendido en cierto modo a ordeñar nuestras mentes y las de los demás; hoy en día, las técnicas de autoestimulación y de estimulación mutua están profundamente enraizadas en nuestra cultura y en nuestra educación. El sentido en que la cultura se ha convertido en un depósito y en un medio de transmisión de innovaciones (no sólo innovaciones de la conciencia) es importante para comprender los orígenes del diseño de la conciencia humana, ya que es otro medio más de evolución.

Uno de los primeros pasos fundamentales que da un cerebro humano durante el impresionante proceso postnatal de autodiseño es el de ajustarse a las condiciones locales que son más importantes: rápidamente (en dos o

13. Este apartado se basa en gran medida en mi artículo «Memes and the exploitation of imagination» (1990a).

tres años), se convierte en un cerebro swahili, japonés o inglés. ¡Vaya paso! ¡Como si para darlo nos lanzaran con una honda!

Para nuestros propósitos no importa si este proceso se denomina aprendizaje o desarrollo diferencial; se produce de forma tan rápida y sin apenas esfuerzo que no cabe duda de que el genotipo humano incluye muchas adaptaciones que están ahí precisamente para facilitar la adquisición del lenguaje. Todo eso se ha producido con gran rapidez, en términos evolutivos, pero eso es precisamente lo que debemos esperar, dado el efecto Baldwin. Ser capaz de hablar es un buen truco tan bueno que cualquiera que tardara demasiado en adquirirlo se hallaría en notable desventaja. Es casi seguro que al primero en hablar de nuestros antepasados le costó mucho más trabajo coger el tranquillo, pero nosotros somos descendientes de los más virtuosos del grupo.[14]

Una vez que los cerebros han abierto las vías de entrada y salida para los vehículos del lenguaje, enseguida se ven atacados por *parásitos* (en el sentido literal del término, como veremos), por unas entidades que han evolucionado precisamente para medrar en ese espacio: los *memas*. Las ideas generales de la teoría de la evolución por selección natural son claras: la evolución se produce siempre que se dan las siguientes condiciones:

1. variación: una abundancia continua de elementos diferentes;
2. herencia o replicación: los elementos tienen la capacidad de crear copias o réplicas de sí mismos;
3. «idoneidad» diferencial: el número de copias de un elemento que se crean en un momento determinado varía, dependiendo de interacciones entre las características del elemento (sea lo que sea lo que lo hace diferente de los demás elementos) y las características del entorno en que persiste.

Nótese que esta definición, aunque tomada de la biología, no dice nada que sea específico de las moléculas orgánicas, la nutrición e incluso la vida. Es una caracterización bastante más general y abstracta de la evolución por selección natural. Como ha señalado el zoólogo Richard Dawkins, el principio fundamental es que

> ...toda vida evoluciona por la supervivencia diferencial de entidades replicadoras. El gen, la molécula de ADN, sucede que es la entidad replicadora que prevalece en nuestro propio planeta. Puede haber otras. Si las hay, siempre que se den otras condiciones, tenderán, casi inevitablemente, a convertirse en la base de un proceso evolutivo.
>
> Pero, ¿debemos trasladarnos a mundos distantes para encontrar otros tipos de replicadores y, por consiguiente, otros tipos de evolución? Pienso que

14. Para un excelente y reciente repaso de todas las controversias suscitadas en la bibliografía sobre la evolución del lenguaje, véase Pinker y Bloom (1990), y los comentarios que siguen.

> un nuevo tipo de replicador ha surgido recientemente en este mismo planeta. Lo tenemos frente a nuestro rostro. Se encuentra todavía en su infancia, aún flotando torpemente en su caldo primario, pero ya está alcanzando un cambio evolutivo a una velocidad que deja al antiguo gen jadeante y muy atrás (1976, pág. 206).*

Estos nuevos replicadores son, más o menos, las ideas. No las «ideas simples» de Locke y Hume (la idea de rojo, la idea de redondo o caliente o frío), sino el tipo de ideas complejas que se constituyen en unidades memorables distintas; son ideas tales como:

la rueda
ir vestidos
la vendetta
el triángulo rectángulo
el alfabeto
el calendario
la *Odisea*
el cálculo
el ajedrez
el dibujo en perspectiva
la evolución por selección natural
el Impresionismo
«Greensleeves»
el desconstruccionismo

Intuitivamente, éstas son unidades culturales más o menos identificables, pero podemos precisar un poco más sobre cómo trazamos los límites, sobre por qué la secuencia re-fa sostenido-la no es una unidad, mientras que el tema del movimiento lento de la Séptima sinfonía de Beethoven sí lo es: las unidades son los elementos mínimos capaces de replicarse con mayor fidelidad y fecundidad. Dawkins acuña un término para referirse a dichas unidades: *memas*:

> ...una unidad de transmisión cultural, o una unidad de imitación. [Mimema] se deriva de una apropiada raíz griega, pero deseo un monosílabo que suene algo parecido a [gen]. (...) como alternativa... se relaciona con [memoria] o con la palabra francesa *même*. (...) Ejemplos de memas son: tonadas o sones, ideas, consignas, modas en cuanto a vestimenta, formas de fabricar vasijas o de construir arcos. Al igual que los genes se propagan en un acervo génico al saltar de un cuerpo a otro mediante los espermatozoides o los óvulos, así los memas se propagan en el acervo de memas al saltar de un cerebro a otro mediante un proceso que, considerado en su sentido más amplio, puede llamarse de imitación. Si un científico escucha o lee una buena idea, la transmite a sus cole-

* Página 251 de la edición castellana (Salvat, 1993). [N. del T.]

gas y estudiantes. La menciona en sus artículos y ponencias. Si la idea se hace popular, puede decirse que se ha propagado, esparciéndose de cerebro en cerebro (1976, pág. 206).*

En *El gen egoísta*, Dawkins nos invita a tomarnos la idea de la evolución de los memas de forma literal. La evolución de los memas no es simplemente algo análogo a la evolución biológica o genética, ni un proceso que pueda ser descrito metafóricamente en términos evolucionistas; es un fenómeno que obedece sin excepción a todas las leyes de la selección natural. La teoría de la evolución por selección natural es neutral en lo que concierne a las diferencias entre genes y memas; se trata simplemente de tipos diferentes de replicadores evolucionando en medios distintos y a ritmos distintos. Así, del mismo modo que los genes de los animales no pudieron existir en este planeta hasta que la evolución de las plantas les abrió el camino (creando una atmósfera rica en oxígeno y proporcionándoles una fuente de nutrientes convertibles), la evolución de los memas no pudo iniciarse hasta que la evolución de los animales abrió el camino, creando una especie, el *Homo sapiens*, con un cerebro que pudiera proporcionar cobijo a los memas y unos hábitos de comunicación que servirían de medio de transmisión de los mismos.

Ésta es una nueva manera de pensar sobre las ideas. También es, como espero demostrar, una buena manera de hacerlo, aunque al principio la perspectiva que nos propone es un tanto inquietante, incluso detestable. Se puede resumir en el siguiente eslogan:

«Un investigador no es más que la manera que tiene una biblioteca de crear otra biblioteca».

No sé lo que pensará usted, pero en principio a mí no me atrae mucho la idea de que mi cerebro sea una especie de estercolero en el que las larvas de las ideas de otros se renuevan, antes de enviar copias de sí mismas en una especie de diáspora informacional. Es como si le robaran a mi mente su importancia en tanto que autor y crítico. De acuerdo con esta concepción, ¿quién es el que manda, nosotros o nuestros memas?

Evidentemente, no es fácil responder a esta pregunta, y es este hecho el que se halla en el centro de todas las confusiones que rodean a la idea del

* Página 251 de la edición castellana (Salvat, 1993). En esta ocasión, no soy totalmente fiel al texto de la traducción pues estoy en total desacuerdo con la manera en que se ha traducido la voz «meme» del inglés. Los traductores de *El gen egoísta* traducen «meme» (en inglés pronunciado «mi:m») por «meme» en castellano. Sin embargo, el castellano posee el sufijo *-ema* propio de palabras cuyo significado es «unidad lingüística en el plano de una descripción idealizada»; es el que encontramos en palabras como fonema o morfema. Me parece, por tanto, más apropiado traducir «meme» por mema, generalizando el significado del sufijo *-ema* a «unidad mínima en el plano de una descripción idealizada»; por otra parte, así respetamos el patrón de traducción ya establecido, ya que fonema y morfema se corresponden con las palabras inglesas «phoneme» y «morpheme», respectivamente. [N. del T.]

yo. La conciencia humana es en gran medida no sólo el producto de la selección natural, sino también de la evolución cultural. La mejor manera de ver la contribución de los memas a la creación de nuestras mentes es seguir muy de cerca los pasos típicos de todo razonamiento evolucionista.

La primera regla para los memas, así como para los genes, es que la replicación no es necesaria por ningún objetivo en particular ni en beneficio de nada ni nadie; los replicadores que triunfan son los que son buenos en el acto de... replicarse, sin importar el motivo por el cual lo hacen. En palabras de Dawkins:

> Un mema que hiciera que su cuerpo se tirara por un barranco correría la misma suerte que un gen que hiciera que los cuerpos se tiren por un barranco. Tendería a ser eliminado del acervo de memas. (...) Sin embargo, ello no significa que el criterio último para la selección de los memas sea la supervivencia de los genes. (...) Evidentemente, un mema que provoque el suicidio de los individuos que lo transportan tiene una gran desventaja, aunque no necesariamente una desventaja fatal (...) un mema suicida puede extenderse, como ocurre cuando un martirio bien publicitado inspira a otros a morir por alguna causa muy querida, lo que a su vez inspira a otros a morir y así sucesivamente (1982, págs. 110-111).

El punto importante aquí es que no existe ninguna conexión *necesaria* entre el poder replicativo del mema, su «idoneidad» desde su punto de vista, y su contribución a *nuestra* idoneidad (sean cuales sean los criterios que utilicemos para evaluarla). La situación no es tan desesperada. Mientras algunos memas nos manipulan claramente para colaborar en su replicación *a pesar de* que los consideremos inútiles e incluso peligrosos para nuestra salud y nuestro bienestar, muchos —la mayoría de ellos si somos afortunados— de los memas que se replican lo hacen no sólo con nuestra bendición, sino también *a causa de* la estima que sentimos por ellos. No creo que haya mucho desacuerdo en el hecho de que, teniéndolo todo en cuenta, algunos memas son buenos desde *nuestro punto de vista*, y no sólo desde su propio punto de vista en tanto que autorreplicadores egoístas: memas generales tales como la cooperación, la música, la escritura, la educación, la preocupación por el entorno, el desarme; y memas particulares tales como *Las bodas de Fígaro, Moby Dick*, los envases retornables, los acuerdos SALT. Otros memas son más discutibles; podemos comprender por qué se extienden y por qué, teniéndolo todo en cuenta, debemos tolerarlos a pesar de los problemas que nos causan: los grandes centros comerciales, la *fast food*, los anuncios en televisión. Finalmente, otros son incuestionablemente perniciosos, pero muy difíciles de erradicar: el antisemitismo, los secuestros de aviones, los virus de ordenador, las pintadas en las paredes.

Los genes son invisibles; son transportados por vehículos para genes (los organismos) en los que tienden a producir los efectos característicos (los efectos «fenotípicos») que determinan, a largo plazo, sus destinos. Los memas también son invisibles, y son transportados por vehículos para memas, las

imágenes, los libros, los dichos (en lenguas determinadas, orales o escritos, sobre papel o en soporte magnético, etc.). Las herramientas, los edificios y otros inventos también son vehículos para memas. Un carro con ruedas radiadas no sólo lleva grano u otras mercancías de un lugar a otro; lleva la brillante idea de un carro con ruedas radiadas de una mente a otra. La existencia de un mema depende del encarnamiento físico en algún medio; si todos estos encarnamientos físicos se destruyen, el mema se destruye. Evidentemente, puede hacer una segunda aparición de forma independiente, del mismo modo que los genes de un dinosaurio podrían, en principio, volver a aparecer juntos en algún futuro distante, aunque los dinosaurios que crearían y habitarían en la Tierra no serían descendientes de los dinosaurios originales; en todo caso serían tan descendientes directos como lo somos nosotros ahora. El destino de los memas —la posibilidad de que copias y copias de las copias de los mismos persistan y se multipliquen— depende de las fuerzas selectivas que actúan directamente sobre los vehículos físicos que los encarnan.

Los vehículos para memas habitan nuestro mundo junto a toda la fauna y la flora, grande y pequeña. Sin embargo, son en gran medida «visibles» sólo para la especie humana. Considérese el entorno de una paloma neoyorquina típica, cuyos ojos y oídos son asaltados cada día por casi tantas palabras, imágenes y otros signos y símbolos que asaltan a cualquier neoyorquino humano. Estos vehículos para memas físicos pueden afectar de manera importante al bienestar de la paloma, pero no en virtud de los memas que transportan; para la paloma no es importante si encuentra unas migajas de pan debajo de una página del *National Enquirer* o del *New York Times*.

Para los seres humanos, en cambio, cada vehículo para memas es un amigo o un enemigo en potencia, portador de un regalo que mejorará nuestros poderes o de un Caballo de Troya que nos distraerá, abrumará nuestra memoria, y desquiciará nuestro juicio. Podemos comparar estos invasores aéreo-transportados de nuestros ojos y oídos con los parásitos que entran en nuestros cuerpos por otras rutas: existen parásitos beneficiosos como las bacterias de nuestro aparato digestivo sin las cuales no podríamos digerir los alimentos, parásitos tolerables que no vale la pena tomarse la molestia de eliminar (como los habitantes de nuestra piel y nuestro cuero cabelludo, por ejemplo), e invasores perniciosos que son muy difíciles de erradicar (como el virus del SIDA, por ejemplo).

Hasta ahora la perspectiva del ojo-del-mema puede todavía parecer una simple manera gráfica de organizar las ya familiares observaciones sobre cómo los elementos de nuestras culturas nos afectan y se afectan entre sí. Pero Dawkins sugiere que nuestras explicaciones tienden a pasar por alto el hecho fundamental de que «una característica cultural puede haber evolucionado de la manera que lo ha hecho simplemente porque es *ventajoso para ella misma*» (1976, pág. 214).* Ésta es la clave para responder a la pregunta

* Página 260 de la traducción castellana (Salvat, 1993). [N. del T.]

de si el mema de los memas es uno de los que merece ser explotado y replicado o no. De acuerdo con la visión tradicional, las siguientes frases eran casi tautológicas:

Las personas creían la idea *X* porque *X* era considerada verdadera.
Las personas tenían en buen concepto *X* porque encontraban *X* hermoso.

Lo que requiere una explicación especial son aquellos casos en que, a pesar de la verdad o la belleza de la idea, ésta *no* es aceptada, o cuando, a pesar de su fealdad o su falsedad, *lo es*. La visión del ojo-del-mema pretende ser una perspectiva general alternativa desde la cual este tipo de desviaciones pueden ser explicadas; para ella, lo que es tautológico es:

El mema *X* se extendió entre las personas porque *X* era un buen replicador.

Ahora bien, existe una correlación no aleatoria entre ambas perspectivas; y no es accidental. No sobreviviríamos si solamente dispusiéramos de un hábito de escoger al azar los memas que pueden ayudarnos. Nuestros sistemas inmunológicos para memas no son infalibles, pero tampoco inútiles. Como regla general, podemos confiar en la coincidencia de ambas perspectivas: en gran medida, los buenos memas son también los que son buenos replicadores.

La teoría gana en interés únicamente cuando nos detenemos a considerar las excepciones, las circunstancias en las que se produce el divorcio de ambas perspectivas; sólo si la teoría de los memas nos permite comprender mejor las desviaciones del esquema normal, tendrá alguna garantía de ser aceptada. (Nótese que, en sus propios términos, el hecho de que el mema de los memas llegue a replicarse con éxito es totalmente independiente de sus virtudes epistemológicas; podría extenderse a pesar de ser pernicioso, o podría extinguirse a pesar de sus virtudes.)

Hoy en día los memas se extienden por el mundo a la velocidad de la luz, y se replican a una velocidad tal, que, en comparación, las moscas de la fruta o las células de la levadura nos parecerían congeladas. Saltan promiscuamente de vehículo en vehículo, y de medio en medio, y se están mostrando prácticamente imposibles de poner en cuarentena. Los memas, como los genes, son *potencialmente* inmortales, pero, también como los genes, dependen de la existencia de una cadena continuada de vehículos físicos, persistentes ante la segunda ley de la termodinámica. Los libros son relativamente permanentes, y las inscripciones en los monumentos aún lo son más, pero, a menos que ambos se hallen bajo la protección de unos conservadores humanos, tienden a desaparecer con el tiempo. Como ocurre con los genes, la inmortalidad es más cuestión de replicarse que de longevidad de los vehículos. La conservación de los memas platónicos, a través de una serie de copias de copias, es un caso típico y sorprendente. Aunque se han descubierto

recientemente algunos fragmentos de papiros con textos de Platón, presuntamente contemporáneos del filósofo, la supervivencia de los memas apenas debe nada a una persistencia tan larga como ésta. Nuestras bibliotecas contienen miles, si no millones, de copias (y traducciones) del *Menón* de Platón, pero los antepasados clave en la transmisión de este texto quedaron reducidos a polvo hace ya muchos siglos.

La mera replicación de los vehículos no es suficiente para asegurar la longevidad de los memas. Varios miles de copias de un nuevo libro encuadernadas en cartoné pueden desaparecer en unos pocos años, y quién sabe cuántas brillantes cartas al director, reproducidas en cientos de miles de copias, desaparecen cada día en sumideros e incineradores. Puede que llegue un día en que sea suficiente con unos evaluadores de memas no humanos para seleccionar unos memas determinados y asegurar así su conservación, pero, por el momento, los memas todavía dependen, aunque sea indirectamente, de que uno o más de sus vehículos pase un cierto tiempo, en forma de crisálida, en un nido de memas muy especial: una mente humana.

Las mentes son un bien escaso, y cada mente tiene una capacidad limitada para albergar memas y, por tanto, hay una considerable competición entre los memas por entrar en tantas mentes como les sea posible. Esta competición es la fuerza selectiva principal de la memosfera, e, igual que en la biosfera, se ha hecho frente a este desafío con grandes dosis de ingenio. Por ejemplo, cualesquiera que sean las virtudes (desde nuestro punto de vista) de los memas siguientes, todos ellos tienen en común la propiedad de poseer expresiones fenotípicas que tienden a aumentar las posibilidades de su propia replicación al debilitar o al contrarrestar las fuerzas del entorno que tenderían a facilitar su extinción: el mema de la *fe*, que rechaza el ejercicio de ese tipo de juicio crítico que llevaría a la conclusión de que la idea de la fe es a todas luces una idea peligrosa (Dawkins, 1976, pág. 212);* el mema de la *tolerancia* o el de la *libertad de expresión*; el mema de incluir en una carta, dentro de una de esas cadenas de cartas que todos hemos recibido, la advertencia de que la ruptura de la cadena ha traído desgracias terribles a todos aquellos que la rompieron con anterioridad; el mema de la *teoría de la conspiración*, que lleva incorporada la respuesta a la objeción de que no existe ningún indicio de que exista tal conspiración: «¡Claro que no, eso es precisamente lo que prueba que la conspiración es muy poderosa!». Algunos de estos memas son «buenos»; otros, quizá, son «malos»; lo que tienen en común es un efecto fenotípico que tiende a desactivar las fuerzas selectivas dirigidas contra ellos. En igualdad de condiciones, la memética de poblaciones predice que los memas de la teoría de la conspiración perdurarán independientemente de si la teoría es cierta o no, y el mema de la fe es apropiado para asegurar su propia supervivencia y la de los memas religiosos que cabalgan con él incluso en los entornos más racionalistas. En efecto, el mema de la fe presenta una *idoneidad dependiente de la frecuencia*:

* Página 259 de la traducción castellana (Salvat, 1993). [N. del T.]

florece cuando se ve superado por los memas racionalistas y, en un entorno con pocos escépticos, tiende a caer en desuso y a desaparecer.

Otros conceptos de la genética de poblaciones se pueden adaptar fácilmente. He aquí un caso de lo que un genetista denominaría *lugares ligados*: dos memas que están físicamente enlazados de modo que siempre tienden a replicarse juntos, lo cual afecta a sus oportunidades. Existe una magnífica marcha ceremonial, conocida por muchos de nosotros, y que se utilizaría mucho más en inauguraciones, bodas y otras ocasiones de corte festivo, amenazando de extinción quizás a «Pompa y circunstancia» y a la «Marcha Nupcial» de *Lohengrin*, si no fuera porque su mema musical está demasiado ligado al mema de su título, en el que muchos de nosotros tendemos a pensar tan pronto como escuchamos la música: la obra maestra de sir Arthur Sullivan «Behold the Lord High Executioner».*

Todos los memas dependen de la posibilidad de alcanzar el refugio de una mente humana, pero una mente humana es también un artefacto creado cuando los memas reestructuran un cerebro humano a fin de convertirlo en un hábitat más apropiado para sí mismos. Las vías de entrada y salida se modifican para adecuarse a las condiciones del entorno y se refuerzan a través de diversos dispositivos artificiales que potencian la fidelidad y la prolijidad de la replicación: las mentes nativas chinas son muy distintas de las mentes nativas francesas, y las mentes alfabetizadas son distintas de las mentes analfabetas. En contrapartida, lo que los memas aportan a los organismos que los albergan es un incalculable almacén de ventajas, con algún que otro Caballo de Troya incluido, sin duda. Los cerebros humanos normales no son todos iguales; varían considerablemente en cuanto a tamaño, forma y un sin fin de detalles en cuanto a las conexiones de las que depende su destreza. Sin embargo, las diferencias más notables en cuanto a la destreza humana dependen de diferencias microestructurales inducidas por los distintos memas que han entrado y fijado allí su residencia. Los memas potencian sus oportunidades entre sí: el mema para la educación, por ejemplo, es un mema que refuerza el proceso mismo de la implantación de memas.

Pero si es cierto que las mentes humanas son también en gran medida la creación de unos memas, entonces no podemos seguir manteniendo la perspectiva con que empezamos; ya no puede tratarse de una cuestión de «memas contra nosotros», porque anteriores invasiones de memas ya han jugado un papel fundamental en la determinación de *quiénes somos o qué somos*. La mente «independiente» luchando por protegerse a sí misma de memas extraños y dañinos no es más que un mito; existe, en el trasfondo, una tensión persistente entre el imperativo biológico de los genes y los imperativos de los memas, pero sería estúpido el «alinearse» con nuestros genes, es decir, el cometer el más craso error de la sociobiología pop. ¿A qué podemos

* El título de la marcha podría traducirse aproximadamente como «Gloria al Señor, Gran Ejecutor», aunque presumiblemente los problemas con el título a que se refiere el autor se deban al hecho de que la palabra *executioner* («ejecutor») significa también «verdugo». [N. del T.]

asirnos, entonces, mientras luchamos por no perder pie en la tormenta de memas en que nos hallamos sumidos? Si el poder replicativo no hace justicia, ¿cuál será el ideal eterno en relación al que «nosotros» juzgaremos el valor de los memas? Es preciso señalar que los memas para conceptos normativos —para el *deber*, el *bien*, la *verdad* y la *belleza*— se hallan entre los más arraigados habitantes de nuestras mentes, y de entre los memas que nos constituyen, son los que juegan un papel más fundamental. Nuestra existencia en tanto que nosotros, como lo que somos en tanto que entes pensantes —y no como lo que somos en tanto que organismos— no es independiente de estos memas.

En resumidas cuentas, la evolución de los memas posee el potencial de contribuir con sustanciales mejoras de diseño a los mecanismos subyacentes del cerebro, con gran rapidez, por comparación al paso lento de la I+D genética. La desacreditada idea lamarckiana de la transmisión genética de las características individuales adquiridas resultó, en un primer momento, atractiva para los biólogos en parte por su presunta capacidad de acelerar la inclusión de nuevas invenciones en el genoma. (Para una excelente y demoledora refutación del lamarckismo, véase la discusión de Dawkins en *The Extended Phenotype*, 1982.) Esto no ocurrirá ni puede ocurrir. El efecto Baldwin acelera la evolución, favoreciendo el movimiento de buenos trucos descubiertos individualmente hacia el genoma, por la vía indirecta de crear nuevas presiones selectivas resultantes de la amplia adopción de esos buenos trucos por parte de los individuos. Sin embargo, la evolución cultural, que se produce aún más deprisa, permite que los individuos adquieran, a través de la transmisión cultural, buenos trucos perfeccionados por predecesores que ni siquiera son sus antepasados genéticos. Tan poderosos son los efectos de dicha posibilidad de compartir buenos diseños que la evolución cultural probablemente ha acabado con casi todas las suaves presiones del efecto Baldwin. Es probable que las mejoras en el diseño que uno recibe de la propia cultura —uno casi nunca tiene que «reinventar la rueda»— limen gran parte de las diferencias genéticas en el diseño del cerebro, eliminando así las ventajas de aquellos que eran un poco mejores en el momento de nacer.

Los tres medios —la evolución genética, la plasticidad fenotípica y la evolución memética— han contribuido sucesivamente al diseño de la conciencia humana a velocidades cada vez mayores. Comparada con la plasticidad fenotípica, que existe desde hace varios millones de años, la evolución memética significativa es un fenómeno extremadamente reciente, convertido en una fuerza poderosa sólo en los últimos cien mil años, y que ha estallado con el desarrollo de la civilización hace menos de diez mil años atrás. Está restringida a una única especie, el *Homo sapiens*, y observamos que con ella nos estamos acercando a un potencial cuarto medio de I+D, gracias a los memas de la ciencia: el examen directo de los sistemas nerviosos individuales por parte de la ingeniería neurocientífica y el examen del genoma por parte de la ingeniería genética.

7. Los memas de la conciencia: la máquina virtual a instalar

> Aunque un órgano puede no haberse formado originalmente con algún fin específico, si ahora sirve a este propósito está plenamente justificada nuestra afirmación de que ha sido especialmente ideado para ello. De acuerdo con el mismo principio, si un hombre construyera una máquina con algún fin específico, pero utilizara viejos engranajes, resortes y poleas, con sólo alguna pequeña alteración, podría considerarse que la máquina, con todas sus partes, fue especialmente ideada con ese propósito. Así, en la naturaleza, casi cada parte de todo ser viviente ha servido probablemente, con alguna ligera modificación, para diversos propósitos, y ha actuado en la maquinaria viva de muchas y distintas formas anteriores.
>
> CHARLES DARWIN, *El origen del hombre*, 1874

> Un gran cerebro, como un gran gobierno, puede no ser capaz de hacer cosas fáciles de una manera simple.
>
> DONALD HEBB, 1958

> El instinto más poderoso en el desarrollo del hombre es el placer que éste siente por su propia destreza. Le gusta hacer bien lo que hace, y, después de hacerlo bien, le gusta hacerlo mejor.
>
> JACOB BRONOWSKI, 1973

Una de las características de mi historia especulativa es el hacer hincapié en el hecho de que nuestros antepasados, igual que nosotros, sentían placer por los diversos modos de autoexploración relativamente indirectos, estimulándose a sí mismos una y otra vez para ver lo que ocurría. A causa de la plasticidad del cerebro, junto a esa inquietud y curiosidad innatas que nos hacen explorar cada rincón y cada grieta de nuestro entorno (del cual nuestros cuerpos son un elemento importante y omnipresente), no debe sorprendernos, en retrospectiva, que hayamos dado con estrategias de autoestimulación y automanipulación que llevaron a la inculcación de hábitos y disposiciones que alteraron radicalmente la estructura comunicativa interna de nuestros cerebros, ni que estos descubrimientos se convirtieran en parte de nuestra cultura —los memas— y estuvieran disponibles para todos.

La transformación de un cerebro humano a causa de una invasión de memas constituye un cambio radical en la competencia del órgano. Como ya hemos señalado, las diferencias en un cerebro cuya lengua materna es el chino en vez del inglés darían cuenta de las grandes diferencias que se observan en la competencia de ese cerebro, que son reconocibles al instante por la conducta externa, y que se revelan en muchos contextos experimentales. Recuérdese, por ejemplo, la importancia que tiene para el investigador (el heterofenomenólogo), en experimentos con sujetos humanos, el saber si los sujetos han comprendido las instrucciones. Estas diferencias funcionales, aunque es posible que estén físicamente encarnadas en patrones microscópicos de variación en el cerebro, son tan buenas como invisibles para los

investigadores del cerebro, ahora y probablemente para siempre. Así pues, si queremos tener un mínimo acceso a la arquitectura funcional *creada* por dichas invasiones de memas, tendremos que buscar un nivel más alto en el que describirla. Afortunadamente disponemos de ese nivel de descripción, que tomamos prestado de la informática y las ciencias de la computación. El nivel de descripción y de explicación que necesitamos es *análogo* (pero no idéntico) a uno de los «niveles de descripción del *software*» de los ordenadores: lo que necesitamos comprender es de qué manera la conciencia humana puede realizarse en el funcionamiento de una *máquina virtual* creada por los memas del cerebro.

Ésta es la hipótesis que voy a defender:

> La conciencia humana es por sí misma un enorme complejo de memas (o, para ser exactos, de efectos de memas en el cerebro) cuyo funcionamiento debe ser equiparado al de una máquina virtual *«von neumanniana» implementada* en la *arquitectura paralela* del cerebro, la cual no fue diseñada para este tipo de actividades. La potencia de dicha *máquina* virtual se ve enormemente potenciada por los poderes subyacentes del *hardware* orgánico sobre el que corre; sin embargo, al mismo tiempo, muchas de sus características más curiosas y, especialmente, sus limitaciones, pueden explicarse como subproductos de los *kludges** que hacen posible esta curiosa pero efectiva reutilización de un órgano que ya existía con nuevos fines.

De esta maraña terminológica que he utilizado para expresarla, pronto irá tomando forma esta hipótesis. ¿Por qué he utilizado esta jerga? Porque éstos son términos que hacen referencia a conceptos útiles que están al alcance de los que reflexionan sobre la mente desde hace muy poco tiempo. No hay otras palabras capaces de expresar estos conceptos con la misma claridad y, por eso, merece la pena conocerlas. Así pues, con la ayuda de una breve digresión histórica, iré introduciéndolos uno a uno y situándolos en el contexto en el que los vamos a utilizar.

Dos de los más importantes inventores del ordenador fueron el matemático inglés Alan Turing y el físico y matemático húngaro-americano John von Neumann. Aunque Turing adquirió una enorme experiencia práctica diseñando y construyendo las máquinas electrónicas especializadas en descifrar códigos secretos que ayudaron a los aliados a ganar la Segunda Guerra Mundial, fue su trabajo puramente abstracto y teórico al desarrollar el concepto de *Máquina de Turing Universal* lo que abrió el camino para la era de los ordenadores. Von Neumann supo ver cómo usar la abstracción de Turing (que, de hecho, era «filosófica», un experimento mental y no una propuesta en el campo de la ingeniería) para elaborar algo lo bastante concreto como para convertirlo en el diseño (todavía bastante abstracto) de un ordenador electrónico real y utilizable en la práctica. Este diseño abstracto, conocido

* Véase más adelante para una definición del término y para una justificación de que no lo traduzcamos. [N. del T.]

como la *arquitectura de von Neumann*, es el que encontramos prácticamente en cualquier ordenador de los que se construyen hoy en día, desde los gigantescos «ordenadores centrales» al *chip* instalado en el más modesto de los ordenadores personales.

Todo ordenador tiene una arquitectura *fija* o *preconfigurada*, pero dispone también de un elevado grado de plasticidad gracias a la memoria, que puede almacenar tanto programas (lo que también se denomina *software*) como *datos*, esos patrones transitorios elaborados para seguir la pista de todo aquello que se debe representar. Los ordenadores, como los cerebros, no están completamente diseñados desde su nacimiento, sino que poseen una flexibilidad que puede ser utilizada como medio para crear arquitecturas más específicas y disciplinadas, máquinas con fines determinados, cada una con su propia idiosincrasia en el momento de recibir los estímulos provenientes del entorno (a través del teclado u otros dispositivos de entrada) y, si el caso lo requiere, en el momento de producir respuestas (a través de la pantalla de TRC u otros dispositivos de salida).

Estas estructuras temporales que están «hechas de reglas y no de cables» son lo que los informáticos denominan *máquinas virtuales*.[15] Una máquina virtual es lo que usted obtiene cuando impone un determinado patrón de reglas (más literalmente: disposiciones o regularidades de transición) sobre toda esa plasticidad. Considérese a alguien que se ha roto un brazo y lo lleva enyesado. El yeso le estorba los movimientos del brazo y, además, su forma y su peso le obligan a ajustar todos los movimientos del cuerpo. Considérese ahora un mimo (Marcel Marceau, por ejemplo) imitando a alguien con un brazo enyesado; si el mimo es bueno, sus movimientos corporales mostrarán las mismas restricciones: el mimo lleva un *yeso virtual* en el brazo, que es «casi visible». Todo aquel que tenga alguna familiaridad con un procesador de textos conoce como mínimo una máquina virtual, y si ha utilizado más de un procesador de textos o una hoja de cálculo o ha jugado con un juego en el mismo ordenador que utiliza con el procesador de textos, entonces conoce varias máquinas virtuales, turnándose en su existencia dentro de una única máquina real. Las diferencias suelen hacerse muy visibles, a fin de que el usuario sepa con que máquina virtual está interactuando en cada momento.

Todo el mundo sabe que cada programa dota al ordenador de poderes distintos, pero no todo el mundo conoce los detalles. Algunos de estos detalles son importantes para nuestra historia, así que debo solicitar su indulgencia porque voy a hacer un breve y elemental repaso del proceso que inventó Alan Turing.

15. Los puristas podrían objetar que mi uso del término *máquina virtual* tiene un significado un poco más amplio de lo que recomienda el uso en ciencias de la computación e informática. Mi respuesta no puede ser más que, como la Madre Naturaleza, cuando tengo algo útil a mano que puedo «pillar» y utilizar ampliando sus funciones (Gould, 1980), no tengo ningún reparo en hacerlo.

Turing no estaba intentando inventar el procesador de textos o el videojuego cuando llevó a cabo sus hermosos descubrimientos. Estaba pensando, con plena conciencia de ello y de forma introspectiva, sobre cómo él, un matemático, procedía en el momento de resolver problemas matemáticos o de efectuar *cálculos*; dio el importante paso de intentar descomponer la *secuencia* de sus *actos mentales* en sus componentes primitivos. «¿Qué hago», debió de haberse preguntado, «cuando efectúo un cálculo? Bueno, pues, en primer lugar me pregunto qué regla se aplica, después aplico la regla, y entonces anoto el resultado, miro el resultado y entonces me pregunto qué debo hacer ahora, y...» Turing era un pensador extremadamente bien organizado, pero su flujo de conciencia, como el de usted o el mío o el de James Joyce, sin duda era un abigarrado revoltijo de imágenes, decisiones, sospechas, recordatorios, etc., a partir del cual él fue capaz de destilar la esencia matemática: el esqueleto, la secuencia mínima de operaciones capaz de alcanzar el mismo objetivo que obtenía con las floridas y sinuosas actividades de su mente consciente. El resultado fue las especificaciones de lo que hoy conocemos con el nombre de máquina de Turing, una brillante idealización y simplificación de un fenómeno hiperracional e hiperintelectual: un matemático efectuando un cálculo riguroso. La idea básica constaba de cinco componentes:

1. un proceso *serial* (los acontecimientos se producen uno después de otro), en
2. un *espacio de trabajo* muy restringido, en el que
3. se inscriben *datos* e *instrucciones*,
4. procedentes de una *memoria* inerte pero extremadamente fiable,
5. para ser manipulados mediante un conjunto finito de *operaciones primitivas*.

En la formulación original de Turing, el espacio de trabajo fue concebido como una unidad de control con un cabezal capaz de leer un cuadro por vez en una cinta de papel, a fin de ver si había escrito un uno o un cero. En función de lo que «veía» el cabezal, éste o bien borraba el cero o el uno y escribía un nuevo símbolo en el cuadro, o dejaba tal como estaba el contenido del cuadro. Acto seguido, la unidad de control desplazaba la cinta un cuadro hacia la derecha o hacia la izquierda y volvía a leer lo que allí había escrito; cada una de estas operaciones estaba regida por un conjunto finito de instrucciones preconfiguradas que constituían su *tabla de máquina*. La cinta era la memoria.

El conjunto finito de operaciones primitivas imaginado por Turing (o, si se prefiere, los actos «atómicos para la introspección») fue concebido deliberadamente como algo muy simple, de modo que la posibilidad de que fueran ejecutadas de forma totalmente mecánica quedara fuera de toda duda. Es decir, dados los propósitos matemáticos de Turing, era importante que cada paso en el proceso que estaba estudiando fuera tan simple, tan tonto,

que pudiera ser llevado a cabo por cualquier bobalicón, alguien que pudiera ser sustituido por una máquina: LEER, BORRAR, ESCRIBIR, MOVER UN ESPACIO HACIA LA IZQUIERDA, etc.

Evidentemente, enseguida vio que su especificación ideal podía utilizarse, de forma indirecta, como anteproyecto de una máquina computadora real; y también lo vieron otros, en particular, John von Neumann, quien modificó las ideas básicas de Turing para crear una arquitectura abstracta para el primer ordenador digital real que podía ser construido en la práctica. Conocemos esta arquitectura con el nombre de *máquina de von Neumann.*

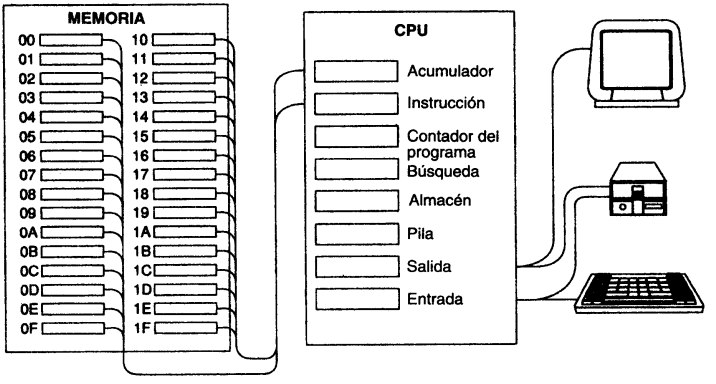


Figura 7.4

A la izquierda está la *memoria* o *memoria de acceso aleatorio* (Random Access Memory o RAM), donde se almacenan datos e instrucciones codificados en forma de secuencias de dígitos binarios o *bits* tales como 00011011 o 01001110. El proceso serial descrito por Turing se lleva a cabo en el espacio de trabajo consistente en dos «registros» señalados como *acumulador* e *instrucción.* Una instrucción se copia electrónicamente en el registro de instrucciones que, acto seguido, la *ejecuta.* Por ejemplo, si la instrucción (traducida al castellano) dice «borrar el acumulador», el ordenador pone el número 0 en el acumulador, y si la instrucción dice «sumar el contenido del registro de memoria 07 al contenido del acumulador», el ordenador buscará el número almacenado en el registro de memoria cuya dirección es 07 (el número puede ser cualquier número) y lo sumará al contenido del acumulador. Y así sucesivamente. ¿Cuáles son las operaciones primitivas? Básicamente las operaciones aritméticas de suma, resta, multiplicación y división; la operaciones de manipulación de datos de búsqueda, almacenamiento, salida y entrada; y (el corazón de la «lógica» de los ordenadores) instrucciones *condicionales* tales como «*SI* el número en el acumulador es mayor que cero, entonces ve a la instrucción en el registro 29; *SI NO,* ve a la instrucción

siguiente». Dependiendo del modelo de ordenador, puede haber desde sólo unas dieciséis operaciones primitivas hasta cientos de ellas, todas configuradas en circuitos con fines específicos. Cada operación primitiva está codificada en forma de un patrón binario (por ejemplo, *sumar* podría ser 1011, y *restar* podría ser 1101), y cada vez que una de estas secuencias entra en el registro de instrucciones, ésta se comporta como un número de teléfono que acabáramos de marcar, abriendo mecánicamente las líneas hacia el circuito correcto: el circuito para sumar, el circuito para restar, etc. Los dos registros en los que sólo puede aparecer al mismo tiempo una instrucción y un valor constituyen el conocido «cuello de botella de von Neumann», el lugar por donde tiene que pasar toda actividad del sistema, en fila india y a través de un paso muy estrecho. En un ordenador muy rápido se ejecutan millones de estas instrucciones por segundo, las cuales, encadenadas por millones, producen el aparentemente mágico efecto que percibe el usuario.

Todos los ordenadores digitales son *descendientes directos* de este diseño; y, pese a que se han llevado a cabo numerosas modificaciones y mejoras, como ocurre con todos los vertebrados, los ordenadores también comparten una misma arquitectura subyacente fundamental. Las operaciones básicas, al poseer esa naturaleza esencialmente aritmética, no parecen a primera vista tener mucho que ver con las «operaciones» básicas de un flujo de la conciencia normal —pensar en París, disfrutar del aroma del pan recién salido del horno, preguntarse dónde pasar las próximas vacaciones—, pero eso no preocupaba a Turing y a von Neumann. Lo que les interesaba era que esa secuencia de acciones podía, «en principio», elaborarse hasta incorporar *todo* «pensamiento racional», y quizá también todo «pensamiento irracional». Es una ironía histórica considerable que dicha arquitectura fuera descrita erróneamente por la prensa popular en el momento en que fue creada. Se denominó a estas nuevas y fascinantes máquinas de von Neumann «cerebros electrónicos gigantes», cuando, de hecho, en realidad eran *mentes electrónicas gigantes*, imitaciones electrónicas —radicales simplificaciones— de lo que William James había bautizado con el nombre de flujo de la conciencia, esa sinuosa secuencia de contenidos mentales conscientes de la que James Joyce hizo las más célebres descripciones en sus novelas. La arquitectura del cerebro es, por el contrario, masivamente paralela, con millones de canales de operación activos al mismo tiempo. Lo que debemos comprender es de qué manera un fenómeno serial joyceano (o, como dije anteriormente, «von neumanniano») puede llegar a existir, con todas sus conocidas peculiaridades, en la barahúnda paralela del cerebro.

He aquí una mala idea: nuestros antepasados homínidos necesitaban pensar de una manera más refinada, más lógica, así que la selección natural poco a poco diseñó e instaló una máquina de von Neumann preconfigurada en el hemisferio izquierdo (el «lógico», el «consciente») del córtex humano. Espero que quede claro, después de nuestro relato evolucionista anterior, que, aunque podría ser una posibilidad *lógica*, ésta carece por completo de plausibilidad biológica. Con la misma facilidad, nuestros antepasados po-

drían haber desarrollado alas o haber nacido con pistolas en las manos; no es así como funciona la evolución.

Sabemos que en el cerebro hay algo que se parece, aunque sea muy *remotamente*, a una máquina de von Neumann, porque, por «introspección», sabemos que poseemos mentes conscientes y que las mentes que así descubrimos tienen en común con las máquinas de von Neumann como mínimo esto: ¡ellas fueron las que inspiraron las máquinas de von Neumann! Este hecho histórico ha dejado una huella fósil bastante significativa: cualquier programador de ordenadores le dirá que es endiabladamente difícil programar los ordenadores paralelos que se están desarrollando actualmente, mientras que es relativamente fácil programar una máquina de von Neumann serial. Cuando uno programa una máquina de von Neumann convencional, siempre puede recurrir a un truco bastante útil: cuando las cosas se ponen difíciles, uno se pregunta «¿Qué haría yo si fuera una máquina intentando resolver este problema?», lo cual suele llevar a una respuesta del tipo, «Bien, pues primero haría esto, y luego tendría que hacer esto, etc.». Pero si uno se pregunta «¿Qué haría yo en esta situación si fuera un procesador paralelo de mil canales?», se queda totalmente en blanco; no poseemos ninguna familiaridad personal con procesos que se producen en mil canales a la vez —ni tampoco poseemos ningún «acceso directo a ellos»—, pese a ser precisamente eso lo que está ocurriendo en nuestro cerebro. Nuestro único acceso a lo que se produce en nuestro cerebro se presenta en un «formato» secuencial que posee un sorprendente parecido con la arquitectura de von Neumann, aunque esta formulación invierte el orden histórico.

Como hemos visto, pues, existe una gran diferencia entre la arquitectura serial (estándar) de un ordenador y la arquitectura paralela del cerebro. Este hecho a menudo se cita como una de las principales objeciones a la inteligencia artificial, que intenta crear inteligencia parecida a la humana mediante el diseño de programas que (casi siempre) corren en máquinas de von Neumann. ¿Acaso esta diferencia en cuanto a la arquitectura constituye una diferencia con importancia teórica? En cierto sentido no. Turing demostró —y ésta es quizá su contribución más importante— que su Máquina de Turing Universal puede computar cualquier función que cualquier ordenador, sea cual sea su arquitectura, pueda computar. En efecto, la Máquina de Turing Universal es el perfecto camaleón matemático, capaz de imitar *cualquier* otra máquina computadora y de hacer, mientras dure el período de imitación, *exactamente* lo que esta máquina haría. Todo lo que usted tiene que hacer es proporcionar a la Máquina de Turing Universal una descripción adecuada de la otra máquina para que, como Marcel Marceau (la Máquina de Mímica Universal) dotado de una coreografía explícita, acto seguido ésta proceda a producir una imitación perfecta basada en esa descripción; de forma virtual, se convierte en la otra máquina. Un programa de ordenador puede, por tanto, verse ya como una lista de instrucciones primitivas que es preciso seguir, ya como una descripción de la máquina que es preciso imitar.

¿Puede usted imitar a Marcel Marceau imitando a un borracho imitando

a un bateador de béisbol? Puede que para usted lo más difícil sea mantener la distinción entre los diferentes niveles de imitación, pero para las máquinas de von Neumann esto es natural. Una vez usted dispone de una máquina de von Neumann sobre la que trabajar, entonces puede anidar máquinas virtuales como si fueran cajas chinas. Por ejemplo, primero usted puede convertir su máquina de von Neumann en, pongamos por caso, una máquina Unix (el sistema operativo Unix) y después implementar una máquina Lisp (el lenguaje de programación Lisp) sobre la máquina Unix —además de WordStar, Lotus 123, y toda una serie de máquinas virtuales— y después implementar un ordenador que juegue al ajedrez sobre la máquina Lisp. Cada máquina se reconoce por su *interficie de usuario*, la manera en que aparece en la pantalla del TRC y la manera en que responde a la información de entrada; esta autopresentación a menudo se denomina la *ilusión del usuario*, ya que el usuario no puede decir —ni le importa— cómo está implementada en el *hardware* la máquina virtual que está utilizando. Al usuario no le importa si la máquina virtual está a uno, dos, tres o diez niveles por encima del *hardware*.[16] (Por ejemplo, los usuarios de WordStar pueden reconocer la máquina virtual de WordStar e interactuar con ella cada vez que la encuentran, sin importarles las variaciones en el *hardware* subyacente.)

Así pues, una máquina virtual es un conjunto temporal de regularidades altamente estructuradas impuesto sobre el *hardware* subyacente por un *programa*: una receta estructurada con cientos de miles de instrucciones que dotan al *hardware* de un enorme conjunto de hábitos interconectados y de disposiciones a reaccionar. Si usted se fija en los microdetalles de todas esas instrucciones pasando por el registro de instrucciones, entonces los árboles no le dejarán ver el bosque; si adopta una cierta distancia, podrá ver con facilidad la arquitectura funcional que se forma a partir de todos estos microelementos: consta de *cosas* virtuales

tales como bloques de texto, cursores, borradores, *sprays*, archivos

y *lugares* virtuales

como directorios, menús, pantallas, *shells**

conectados por *caminos* virtuales

16. E incluso podría no ser una máquina *virtual*. Podría ser una máquina real hecha por encargo preconfigurada con un propósito específico, como una máquina Lisp, que es una descendiente directa de las máquinas Lisp *virtuales*, y que ha sido diseñada directamente en sus *chips* de silicio para funcionar con el lenguaje de programación Lisp.

* Un *shell* es precisamente ese espacio virtual que queda entre dos máquinas virtuales, como el estrecho espacio que queda entre dos cajas chinas anidadas. Desde un *shell* podemos «observar» o «utilizar» una máquina virtual desde otra que esté por encima o por debajo de ella. Mantenemos la palabra inglesa «shell», porque es ésta la más comúnmente utilizada por los informáticos hispanohablantes; sólo ocasionalmente se utiliza «concha» en textos escritos. [N. del T.]

tales como «ESCapar al Dos», o *entrar* en el menú de IMPRESIÓN *desde* el menú PRINCIPAL

y que permiten llevar a cabo vastas e interesantes *operaciones* virtuales

tales como buscar una palabra en un archivo, o aumentar de tamaño una ventanilla aparecida en la pantalla.

Si cualquier máquina computadora puede ser imitada por una máquina virtual sobre una máquina de von Neumann, se puede deducir que, si el cerebro es una máquina de procesamiento paralelo masivo, ésta también puede ser imitada sin ningún problema por una máquina de von Neumann. Y desde el principio de la era de los ordenadores, los teóricos han utilizado este poder camaleónico de las máquinas de von Neumann para crear arquitecturas paralelas *virtuales* con el propósito de modelar estructuras parecidas a la estructura cerebral.[17] ¿Cómo podemos hacer que una máquina que hace una sola cosa por vez se convierta en una máquina que hace muchas cosas a la vez? Mediante un proceso muy parecido a hacer punto. Supongamos que el procesador paralelo que queremos simular tiene diez canales. Primero se instruye a la máquina de von Neumann para que ejecute las operaciones que debería ejecutar el primer nodo del primer canal (nodo 1 en el diagrama), guardando el resultado en un *buffer* de memoria, después el nodo 2 y así, sucesivamente, hasta que los diez nodos del primer nivel han avanzado un paso. Entonces la máquina de von Neumann se enfrenta a los efectos de cada uno de los resultados del primer nivel sobre los nodos del

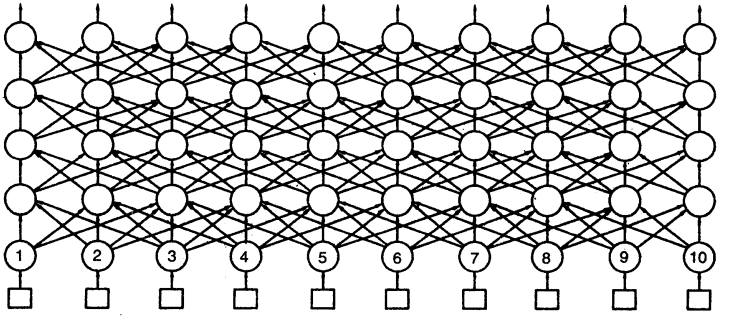


Figura 7.5

17. Las «neuronas lógicas» de McCulloch y Pitts (1943) se desarrollaron contemporáneamente a la invención del ordenador serial e influyeron en el pensamiento del propio von Neumann, y dieron lugar a los perceptrones de los años cincuenta, los antepasados del conexionismo actual. Para una breve reseña histórica de estos temas, véase Papert (1988).

segundo nivel, recuperando del *buffer* de memoria los resultados previamente calculados uno cada vez y aplicándolos como entradas del siguiente nivel. Así procede, laboriosamente, tejiendo hacia adelante y hacia atrás, cambiando tiempo por espacio. Para simular una máquina virtual de diez canales se necesita *como mínimo* diez veces más tiempo que para simular una máquina de un canal, y para simular una máquina de un millón de canales (como podría ser el cerebro) se necesitaría como mínimo un millón de veces más tiempo de simulación. La demostración de Turing no dice nada sobre la velocidad con que se puede llevar a cabo la imitación y, para algunas arquitecturas, incluso la impresionante velocidad de los modernos ordenadores digitales se ve superada por la amplitud de ciertas tareas. Éste es el motivo por el cual los investigadores en IA interesados en estudiar la potencia de las arquitecturas paralelas han desviado su atención hacia las arquitecturas paralelas reales —dispositivos que con mayor justicia merecerían el nombre de «cerebros electrónicos gigantes»— sobre las que construir sus simulaciones. Pero, en principio, cualquier máquina paralela puede ser remedada perfectamente, aunque de forma ineficiente, como una máquina virtual sobre una máquina de von Neumann.[18]

Ahora estamos en disposición de volver del revés esta idea tradicional. Del mismo modo que uno puede simular un cerebro paralelo en una máquina de von Neumann, también puede, en principio, simular (o algo parecido) una máquina de von Neumann sobre un *hardware* paralelo; y esto es precisamente lo que quiero sugerir: las mentes humanas conscientes son máquinas virtuales más o menos seriales implementadas —de forma ineficiente— sobre el *hardware* paralelo que la evolución nos ha legado.

¿Qué cosa juega el papel de «programa» cuando hablamos de una máquina virtual corriendo sobre el *hardware* paralelo del cerebro? Lo importante es que haya grandes cantidades de plasticidad ajustable que puedan encargarse de un sinfín de microhábitos distintos, y así encargarse también de distintos macrohábitos. En el caso de la máquina de von Neumann, esto se consigue gracias a cientos de miles de ceros y unos *(bits)*, divididos en «palabras» de 8, 16, 32 o 64 *bits*, dependiendo de la máquina. Las palabras se almacenan por separado en registros de la memoria y se accede a ellas, una por vez, en el registro de instrucciones. En el caso de una arquitectura paralela podemos conjeturar que esto se consigue gracias a configuraciones de miles, millones o miles de millones de conexiones reforzadas entre neuronas, que, todas a la vez, proporcionan al *hardware* subyacente un nuevo conjunto de macrohábitos, un nuevo conjunto de regularidades condicionales de conducta.

¿Y cómo se instalan estos programas de millones de conexiones reforzadas en el ordenador del cerebro? En una máquina de von Neumann basta con «cargar» el programa de un disco a la memoria principal, y así el orde-

18. Para más detalles sobre las implicaciones de la velocidad real, y las implicaciones para la inteligencia artificial, véase el artículo «Fast thinking» en mi libro *La actitud intencional* (1987a).

nador se ve instantáneamente dotado de todo un conjunto de nuevos hábitos; con los cerebros es preciso pasar por un período de *formación*, incluido, especialmente, el tipo autoestimulación repetitiva que esbozamos en la sección 5. Ésta es, sin duda, una diferencia fundamental. La unidad central de procesamiento (Central Processing Unit o CPU) de una máquina de von Neumann es rígida en el sentido de que responde a las cadenas de *bits* que componen sus palabras, tratándolas como instrucciones de un *lenguaje máquina* que le es propio y que es fijo. Tales propiedades son definitorias de lo que podríamos denominar *ordenador digital con programa almacenado*, y el cerebro humano es algo completamente diferente. Mientras que probablemente sea cierto que cada configuración de conexiones reforzadas entre neuronas en el cerebro tenga un efecto determinado sobre el comportamiento resultante de la red de neuronas que la rodea, no hay ningún motivo para suponer que dos cerebros distintos posean el «mismo sistema» de interconexiones; por este motivo, es casi seguro que no existe nada ni remotamente parecido al lenguaje máquina fijo que, por ejemplo, todos los ordenadores IBM y compatibles comparten. Así pues, si dos o más cerebros «comparten *software*», ello no se deberá a un simple y directo proceso análogo a copiar un programa en lenguaje máquina de una memoria a otra. (Asimismo, es evidente que la plasticidad que de un modo u otro favorece la memoria en el cerebro no es algo aislado como un mero depósito pasivo; la división del trabajo entre memoria y CPU es artificial, y no tiene parangón en el cerebro, un asunto que retomaremos en el capítulo 9.)

Si existen tan importantes —y a menudo olvidadas— diferencias, ¿por qué insisto entonces en establecer un paralelismo entre la conciencia humana y el *software* de un ordenador? Porque, como espero demostrar, ciertas características importantes de la conciencia humana, que en caso contrario seguirían siendo un misterio, son susceptibles de recibir una explicación reveladora, en el supuesto de que (1) la conciencia humana es una innovación demasiado reciente como para estar preconfigurada en la maquinaria innata, (2) es en gran medida un producto de la evolución cultural que se imparte a los cerebros en las primeras fases de su formación, y (3) de que el éxito de su instalación está determinado por un sinfín de micro-disposiciones en la plasticidad del cerebro, lo cual significa que es muy posible que sus rasgos funcionalmente importantes sean invisibles al examen neuroanatómico a pesar de lo destacado de los efectos. Del mismo modo que a ningún informático se le ocurriría intentar comprender las ventajas y los inconvenientes de WordStar frente a WordPerfect a partir de información sobre las diferencias en los patrones de voltaje de la memoria, ningún investigador en ciencia cognitiva debería esperar comprender la conciencia humana partiendo solamente de datos neuroanatómicos. Además, (4) la idea de la *ilusión del usuario* de una máquina virtual es tentadora y terriblemente sugestiva: si la conciencia es una máquina virtual, ¿quién es el usuario con quien funciona la ilusión del usuario? Reconozco que las cosas están tomando un cariz sospechoso, que parece conducirnos inexorablemente de nuevo hacia ese yo

cartesiano interno, sentado ante su estación de trabajo cortical y reaccionando ante la ilusión del usuario que produce el *software* que en ella corre; como veremos existen, no obstante, algunas maneras de rehuir tan fatal desenlace.

Supóngase, de momento, que existe una versión más o menos bien diseñada (depurada) de esta máquina virtual del flujo de la conciencia —la máquina joyceana— en la memosfera. Como hemos visto, ya que no existe un lenguaje máquina común a todos los cerebros, los métodos de transmisión que garantizarían la presencia de una máquina virtual lo bastante uniforme operando a través de la cultura deben ser sociales, altamente dependientes del contexto y, hasta cierto punto, autoorganizativos y autocorrectivos. Hacer que dos ordenadores diferentes —por ejemplo, un Macintosh y un IBM-PC— «hablen el uno con el otro» es una cuestión de intrincada ingeniería, que obliga a ocuparse de detalles nimios y que depende de un conocimiento preciso de la maquinaria interna de ambos sistemas. En tanto que los seres humanos pueden «compartir *software*» sin que nadie posea estos conocimientos, ello debe ser porque los sistemas compartidos poseen un elevado grado de labilidad y de tolerancia a los diversos formatos. Existen diversos métodos para compartir dicho *software*: el aprendizaje por imitación, el aprendizaje como resultado de un «refuerzo» (sea éste impuesto deliberadamente por un profesor —recompensa, ánimo, desaprobación, amenaza— o transmitido sutil e inconscientemente en el curso de encuentros comunicativos), y el aprendizaje como resultado de un proceso de formación explícito, llevado a cabo en una lengua natural previamente aprendida a través de los dos métodos anteriores. (Piénsese, por ejemplo, en el tipo de hábitos que se conformarían al decir frecuentemente a un novicio, «cuéntame qué estás haciendo» y «dime por qué lo estás haciendo». Ahora, piénsese en el novicio adquiriendo el hábito de formularse a sí mismo esas preguntas.)

De hecho, presumo que no sólo el lenguaje hablado sino también el escrito juega un papel fundamental en el desarrollo y la elaboración de las máquinas virtuales que muchos de nosotros hacemos correr la mayor parte del tiempo en nuestros cerebros. Del mismo modo que la rueda es un desarrollo tecnológico cuya utilidad depende en gran medida de la existencia de raíles o vías pavimentadas o cualquier otro tipo de superficie allanada artificialmente, la máquina virtual de que estoy hablando sólo puede existir en un entorno con lenguaje e interacción social, pero también escritura y dibujo, y ello por la sencilla razón de que las exigencias que su implementación impone sobre la memoria y el reconocimiento de patrones requieren que el cerebro «descargue» algunos de sus recuerdos en *buffers* del entorno. (Nótese que ello implica que la «mentalidad ágrafa» podría comportar una clase de arquitecturas virtuales significativamente diferentes de las que encontramos en sociedades que conocen la escritura.)

Piense en el acto de sumar dos números de diez dígitos mentalmente, sin utilizar lápiz y papel ni pronunciar los números en voz alta. Piense en el acto de intentar imaginarse, sin un diagrama, una manera de unir tres autopis-

tas en un cruce en forma de trébol, de modo que, viniendo de cualquier autopista, en cualquiera de las dos direcciones, se pueda entrar en cualquiera de las otras dos autopistas y en cualquiera de las dos direcciones, sin tener que entrar antes en la tercera autopista. Éste es el tipo de problemas que los seres humanos resuelven fácilmente con la ayuda de dispositivos externos de memoria y el uso de sus dispositivos de exploración preexistentes (eso que llamamos ojos y oídos), con sus altamente desarrollados circuitos preconfigurados de reconocimiento de patrones. (Véase el capítulo 14 de McClelland y Rumelhart, 1986, para algunas interesantes observaciones sobre este asunto.)

Instalamos en nuestro cerebro un conjunto, organizado y parcialmente puesto a prueba, de hábitos de mente, como los denomina el político y científico Howard Margolis (1987), en el curso de nuestro primer desarrollo infantil. En el capítulo 9 examinaremos de cerca los detalles más verosímiles de esta arquitectura, aunque quiero avanzar que mi propuesta es que la estructura general de este nuevo conjunto de regularidades es la de un *encadenamiento serial*, en el que primero una «cosa» y después otra «cosa» tienen lugar en (aproximadamente) el mismo «sitio». Este flujo de acontecimientos lo conducen una serie de hábitos aprendidos, de los cuales el hábito de hablar con uno mismo es el ejemplo principal.

Habida cuenta de que esta nueva máquina que se ha creado en nosotros es un complejo de memas muy profusamente replicado, podemos preguntarnos a qué se debe su éxito replicativo. Evidentemente, debemos considerar la posibilidad de que *podría* no ser buena para nada, excepto para replicarse. *Podría* ser un programa virus, que parasita fácilmente los cerebros humanos sin dar a los seres humanos cuyos cerebros ha infestado ninguna ventaja para la competición. En una visión más plausible, podríamos decir que *ciertas características* de la máquina podrían ser parásitos, que existen sólo porque pueden, y porque no es posible —o no vale la pena— deshacerse de ellos. William James pensó que sería absurdo suponer que la cosa más impresionante que conocemos en el universo —la conciencia— es algo artificial, que no juega ningún papel esencial en el modo en que funcionan nuestros cerebros, pero, por muy inverosímil que pueda parecer, no es algo que podamos descartar en principio, por lo que tampoco podemos decir que sea realmente absurdo. Hay a nuestro alrededor muchos datos que evidencian los beneficios que la conciencia aparentemente nos proporciona, y es por ello que, sin lugar a dudas, podemos sentirnos satisfechos por sus múltiples *raisons d'être*; pero podemos también caer en el error de malinterpretar dicha evidencia, si pensamos que todo misterio permanecerá a menos que cada una de sus propiedades tenga —o haya tenido en algún momento— una función (desde *nuestra* perspectiva de «usuarios» de la conciencia) (Harnad, 1982). Hay espacio suficiente para ciertos hechos brutales que carezcan por completo de justificación funcional. Algunas características de la conciencia podrían ser, simplemente, memas egoístas.

Visto desde el lado bueno, sin embargo, ¿para la resolución de qué tipo

de problemas está aparentemente diseñada esta máquina? El psicólogo Julian Jaynes (1976) ha presentado argumentos bastante persuasivos en favor de la idea de que sus capacidades de autoexhortación y autorremembranza son un requisito previo para la aparición del tipo de elaborados turnos de autocontrol a largo plazo sin los cuales la agricultura, proyectar edificios, ni ninguna otra actividad civilizada y civilizadora se podría haber organizado. También parece ser buena para el tipo de autosupervisión que puede impedir que un sistema defectuoso sea víctima de sus propios errores, un tema que ha desarrollado Douglas Hofstadter (1985) desde la perspectiva de la inteligencia artificial. Y el psicólogo Nicholas Humphrey (1976, 1983a, 1986) la ha considerado como el medio a través del cual se puede explotar aquello que podríamos denominar simulaciones sociales, utilizando la introspección para guiar nuestras propias intuiciones sobre lo que otros están pensando y sintiendo.

Subyacente a estos talentos avanzados y especializados está la capacidad básica de resolver el metaproblema de en qué pensar después. Antes, en este capítulo, vimos que cuando un organismo se enfrenta a una crisis (o simplemente a un problema nuevo y difícil), éste ya puede poseer ciertos recursos que le serían de mucha utilidad en esas circunstancias *¡si pudiera hallarlos y utilizarlos a tiempo!* Las respuestas de orientación, según ha conjeturado Odmar Neumann, tienen el valioso efecto de más o menos poner en marcha a todo el mundo, pero la consecución de este despertar global, como vimos, es tanto parte del problema como parte de su solución. Apenas servirá de ayuda si, en el siguiente paso, el cerebro no consigue que todos esos voluntarios lleven a cabo algún tipo de actividad coherente. El problema para el cual las respuestas de orientación se presentaban como una solución era el de tener un acceso global y total a un conjunto de especialistas acostumbrados a ocuparse de sus asuntos. Incluso si, gracias a una arquitectura subyacente de tipo pandemónium, el caos se supera enseguida, dejándolo temporalmente todo a cargo de un especialista (que, quizás, está mejor informado a causa de la competición que acaba de ganar), es evidente que hay por lo menos tantas maneras malas como maneras buenas de resolver estos conflictos. Nada garantiza que el especialista políticamente más efectivo sea «la persona ideal para el trabajo».

Platón supo ver el problema con gran claridad hace dos mil años y halló una hermosa metáfora para exponerlo:

> Atiende, entonces, a ver si es posible que podamos poseer así el conocimiento, sin tenerlo. Es lo mismo que si una persona hubiera cogido aves agrestes, como palomas o cualquier otra clase, y las alimentara en un palomar que hubiese dispuesto en su casa. En cierto sentido podríamos decir que las tiene siempre, precisamente porque las posee. ¿No es así? (...) Pero en otro sentido diríamos que no tiene ninguna, sino que ha adquirido un poder sobre ellas, ya que las tiene al alcance de la mano en un recinto propio, de forma que pueda cogerlas o tenerlas cuando quiera. Puede apresar la que desee y soltarla de nuevo, y le es posible hacer esto cuantas veces le parezca. (...) [T]ambién va-

> mos a colocar ahora un cierto palomar en el alma con toda clase de aves, unas en bandadas separadas de las demás, otras, en pequeños grupos, y unas pocas aisladas, volando al azar entre las otras (*Teeteto*, 197-198a).*

Lo que Platón supo ver es que no basta sólo con tener los pájaros; lo difícil es aprender a hacer que acuda el pájaro adecuado cuando uno lo llama. Gracias al *razonamiento*, siguió argumentando Platón, mejoramos nuestra capacidad de hacer que acudan los pájaros adecuados en el momento preciso. Aprender a razonar es, de hecho, aprender estrategias de recuperación del conocimiento.[19] Y éste es el punto en que los hábitos entran en juego. Ya hemos visto en un esbozo un tanto burdo cómo hábitos de mente tan generales como hablar a uno mismo o dibujar para uno mismo *podrían* hacer llegar los fragmentos adecuados de información hasta la superficie (¿la superficie de qué? Esta cuestión queda aplazada hasta el capítulo 10). Pero unos hábitos de mente más específicos, mejoras y nuevas elaboraciones de maneras específicas de hablar a uno mismo, pueden ampliar nuestras oportunidades todavía más.

El filósofo Gilbert Ryle, en su libro *On Thinking* (1979), publicado póstumamente, decidió que el pensar, ese pensar lento, difícil y ponderado en el que aparentemente está sumido el personaje de la famosa estatua de Rodin, el Pensador, debe, sin lugar a dudas, ser en gran medida un acto de hablar a uno mismo. ¡Sorpresa, sorpresa! ¿No es acaso obvio que eso es precisamente lo que hacemos cuando pensamos? Bueno, pues... sí y no. Es obvio que eso es lo que (a menudo) parece que estamos haciendo; a veces, incluso, podemos comunicar a otros cuáles son las palabras que utilizamos en nuestros soliloquios silenciosos. Pero lo que dista mucho de ser obvio es por qué tiene que ser útil el hablar a uno mismo.

> ¿Qué es lo que está aparentemente haciendo *Le penseur* en sus interioridades cartesianas? O, para parecer más científicos, ¿cómo son los procesos mentales que se están llevando a cabo en la *camera obscura* cartesiana? (...) Es un hecho notorio que algunas de nuestras cogitaciones, aunque no todas, acaban con una solución para nuestros problemas; andábamos a ciegas, pero finalmente supimos ver la luz. Pero si a veces tienen éxito, ¿por qué no siempre? Si acaban tardíamente, ¿por qué no más pronto? Si se producen con dificultad, ¿por qué no más fácilmente? En cualquier caso, ¿por qué funciona? ¿Cómo es posible que funcione? (Ryle, 1979, pág. 65).

Los hábitos de mente fueron diseñados a lo largo de los siglos para ir *marcando* el camino por los senderos de exploración más frecuentados. Como señala Margolis,

* Traducción castellana de A. Vallejo Campos en el volumen V de los *Diálogos*, Madrid, Gredos (Biblioteca Clásica Gredos, 117), 1988, págs. 288-289. [N. del T.]

19. Para una interesante discusión sobre el (aparente) desacuerdo entre dos escuelas de pensamiento en el seno de la inteligencia artificial, razonamiento frente a búsqueda, véase Simon y Kaplan, 1989, págs. 18-19.

> ...incluso un ser humano hoy en día (y, por tanto, *a fortiori*, también un antepasado remoto de los seres humanos contemporáneos) no puede, por lo general, mantener con facilidad su atención ininterrumpidamente sobre un único problema durante más de unas pocas décimas de segundo. Y sin embargo trabajamos en problemas que requieren muchísimo más tiempo. La estrategia a que recurrimos para hacerlo (como puede verse si uno se observa a sí mismo) incluye períodos de meditación seguidos de períodos de recapitulación, en los que nos describimos a nosotros mismos lo que ocurrió durante los períodos de meditación, y que conducen a éste o a aquel resultados intermedios. Todo esto tiene una función clara: al ensayar estos resultados intermedios... los implicamos con la memoria, pues los contenidos inmediatos del flujo de la conciencia se pierden rápidamente a menos que se produzca un ensayo. (...) Gracias al lenguaje, podemos describirnos a nosotros mismos lo que aparentemente ocurrió durante la meditación que condujo a un juicio, podemos producir una versión ensayable del proceso de llegar a un juicio, e implicarla con la memoria a largo plazo y, así, ensayarla efectivamente. (Margolis, 1987, pág. 60.)

Aquí, en los hábitos individuales de autoestimulación, es donde debemos buscar *kludges* (que rima con *stooges*), el término que utilizan los *hackers* informáticos para designar aquellos parches *ad hoc* que introducen en un programa durante el proceso de depuración del mismo para que éste siga funcionando. (La lingüista Barbara Partee criticó una vez una chapuza no muy elegante en un programa de IA para el procesamiento del lenguaje por ser un *odd hack* (algo así como un «parche chapucero») —uno de los mejores juegos de palabras que nunca he escuchado. La Madre Naturaleza está llena de parches chapuceros, y es de esperar que también los encontremos en la adopción idiosincrásica de cada individuo de la máquina virtual.)*

He aquí un ejemplo plausible: dado que la memoria humana no fue innatamente diseñada para ser la superfiable, de rápido acceso, memoria de acceso aleatorio (que toda máquina de von Neumann necesita), cuando los diseñadores (cultural y temporalmente distribuidos) de la máquina virtual von neumanniana se enfrentaron con la tarea de fabricar un sustituto adecuado que funcionara en un cerebro, dieron con varios trucos para potenciar la memoria. Esos trucos básicos son ensayar, ensayar y volver a ensayar, asistidos por rimas y máximas con ritmo y fáciles de recordar. (Las rimas y los ritmos explotan ese vasto potencial del sistema preexistente de análisis auditivo para reconocer patrones de sonidos.) La deliberada y repetida yuxtaposición de elementos entre los cuales se necesitaba establecer un lazo asociativo —de modo que un elemento siempre hará que el cerebro «recuer-

* En este párrafo tan lleno de juegos de palabras, me resisto a traducir más de lo traducible y mantengo aquellos términos, como «hacker» y «hack», que, por lo que he podido comprobar, han sido adoptados directamente en la jerga informática. Un «hacker» es un programador poco elegante, que no duda en utilizar cualquier «hack» para que su programa funcione. En cuanto a la palabra «stooge» (que rima con «kludge») designa en inglés coloquial a un ayudante o un subalterno.

de» al otro— se vio ulteriormente mejorada, podemos suponer, al hacer que las asociaciones fueran lo más ricas posibles, vestidas no sólo con rasgos visuales y auditivos, sino también sacando partido del resto del cuerpo. Ese fruncir el ceño y sostenerse la barbilla de *Le penseur*, y ese rascarse la cabeza, ese murmurar, ese pasear o ese emborronar hojas de papel que constituyen la idiosincrasia del pensar de cada uno, podrían no ser sólo productos secundarios y azarosos del pensamiento consciente, sino también contribuciones funcionales (o los vestigios de contribuciones funcionales previas y más burdas) a la laboriosa tarea de disciplinar el cerebro que era preciso llevar a cabo para convertirlo en una mente madura.

Y en lugar del preciso y sistemático «ciclo de buscar y ejecutar» o «ciclo de las instrucciones» que lleva cada nueva instrucción al registro de instrucciones para ser ejecutada, deberíamos buscar «reglas» de transición no muy bien organizadas, un tanto errabundas y muy poco lógicas, donde esa predilección (casi) innata del cerebro por la «asociación libre» recibe cadenas de asociación más largas a fin de asegurar, más o menos, que se ensayarán las secuencias adecuadas. (En el capítulo 9 consideraremos algunos desarrollos de esta idea dentro de la IA; para otros desarrollos que hacen hincapié en aspectos distintos, véase Margolis, 1987, y Calvin, 1987, 1989a. Véase también Dennett, 1991b.) No es de esperar que la mayoría de esas secuencias que se producen sean *algoritmos* bien establecidos, que ofrezcan garantías de llegar a los resultados deseados, sino solamente incursiones afortunadas en el aviario de Platón.

La analogía con las máquinas virtuales de la informática y las ciencias de la computación nos ofrece una perspectiva útil desde la cual observar el fenómeno de la conciencia humana. Originalmente los ordenadores fueron concebidos como grandes calculadoras, pero ahora, con mucha imaginación, se ha sacado partido de esa capacidad de procesar números para crear nuevas máquinas virtuales, como los videojuegos o los procesadores de textos, en las que apenas se percibe ese procesar de números, y en las que esos nuevos poderes casi parecen mágicos. De forma similar, nuestros cerebros no fueron diseñados (con la excepción de algunos órganos periféricos muy recientes) para procesar textos, pero ahora una gran porción —quizás incluso la parte del león— de las actividades que tienen lugar en los cerebros humanos adultos se dedica a una especie de procesamiento de textos: la producción y la comprensión del habla, así como el ensayo serial y el reajuste de los elementos lingüísticos, o mejor, sus sustitutos neuronales. Estas actividades magnifican y transforman los poderes del *hardware* subyacente de maneras que parecen (desde «fuera») bastante mágicas.

Y sin embargo (estoy seguro de que ustedes quieren hacer esta objeción), ¡todo esto tiene poco o nada que ver con la conciencia! Después de todo, una máquina de von Neumann es completamente inconsciente; ¿por qué el mero hecho de implementarla —o algo parecido a ella: una máquina joyceana— debería hacerla más consciente? Tengo una respuesta: la máquina de von Neumann, al estar configurada así desde el principio, con conexiones de má-

xima eficiencia, nunca tuvo que convertirse en el objeto de sus propios y elaborados sistemas perceptivos. Los trabajos de la máquina joyceana, por otra parte, son tan «visibles» y «audibles» para la primera como lo son cualesquiera de los objetos del mundo exterior para cuya percepción está diseñada, por la simple razón de que son en gran medida los mismos mecanismos perceptivos los que se dirigen hacia ellos.

Ahora esto *parece* un juego de espejos, lo sé. Ciertamente, es antiintuitivo, difícil de digerir, y, en principio, una monstruosidad; justo lo que uno esperaría de una idea capaz de romper con siglos de misterio, controversia y confusión. En los dos capítulos siguientes examinaremos con más detalle —y alguna dosis de escepticismo— la manera en que se puede demostrar que ese aparente juego de espejos es parte legítima de una explicación de la conciencia.

Capítulo 8

CÓMO LAS PALABRAS HACEN COSAS CON NOSOTROS

> Y el lenguaje nace, como la conciencia, de la necesidad, de los apremios del intercambio con los demás hombres.
> KARL MARX y FRIEDRICH ENGELS, *La ideología alemana*, 1846*

> Cabe admitir que la conciencia se ha desarrollado únicamente a impulso de la necesidad de comunicarse.
> FRIEDRICH NIETZSCHE, *La gaya ciencia*, 1882†

> Antes de que mi maestro llegara yo no sabía de mi existir. Vivía en un mundo que era un no mundo. No puedo esperar describir adecuadamente ese tiempo inconsciente, y a la vez consciente, de la nada. (...) Como yo no poseía ninguna capacidad de pensamiento, no podía comparar un estado mental con otro.
> HELEN KELLER, *The world I live in*, 1908

1. Repaso: *¿e pluribus unum?*

En el capítulo 5 expusimos la mala idea, persistente y seductora, del Teatro Cartesiano, donde se presenta un espectáculo de luz y de color ante una audiencia solitaria pero poderosa, el ego o el ejecutor central. Aunque hemos podido comprobar por nosotros mismos la incoherencia de esta idea y hemos identificado una alternativa, el modelo de las Versiones Múltiples, el Teatro Cartesiano seguirá embrujándonos hasta que hayamos anclado firmemente nuestra alternativa en la tierra firme de la ciencia empírica. Iniciamos esta tarea con el capítulo 6 y la continuamos, realizando nuevos progresos, con el capítulo 7. Literalmente, tuvimos que regresar a los principios: los principios de la evolución que guiaron una narración especulativa sobre el proceso gradual del desarrollo del diseño que ha creado nuestra conciencia. Ello nos permitió echar una mirada a la maquinaria de la conciencia desde el interior de la caja negra, entre bastidores, podríamos decir, como homenaje a esta tentadora imagen teatral que estamos intentando derrocar.

En nuestros cerebros hay una colección de circuitos cerebrales ensamblados, que, gracias a una serie de hábitos inculcados en parte por la cultura y en parte por la autoexploración individual, conspiran para producir una

* Traducción castellana de Wenceslao Roces, Barcelona, Crítica.
† Traducción castellana de Pedro González Blanco, Olañeta, Palma de Mallorca, 1979.

máquina virtual más o menos ordenada, más o menos efectiva y más o menos bien diseñada: la *máquina joyceana*. Al aunar todos estos órganos especializados, que evolucionaron independientemente, ante una causa común y dotando, así, al conjunto de unos poderes muy mejorados, la máquina virtual, este *software* del cerebro, lleva a cabo una especie de milagro político interno: crea un *capitán virtual* para la tripulación, sin ascender a ninguno de ellos al rango de dictador vitalicio. ¿Quién está al mando? Primero una coalición y luego otra, en una alternancia que no es caótica gracias a unos buenos metahábitos que tienden a producir secuencias coherentes y resueltas en vez de una interminable y atropellada carrera por el poder.

Ese saber ejecutivo resultante no es más que uno de los poderes tradicionalmente atribuidos al yo, aunque uno de los importantes. William James le rindió homenaje cuando satirizó la idea de la neurona pontificia en alguna parte del cerebro. Sabemos que la descripción del trabajo que efectúa este subsistema jefe es incoherente, pero también sabemos que esas responsabilidades de control y esas decisiones deben estar repartidas de *un modo u otro* por el cerebro. Nosotros *no* somos barcos a la deriva en manos de una tripulación amotinada; nos las arreglamos bastante bien no sólo manteniéndonos lejos de los bancos de arena y otros peligros, sino también planeando campañas, corrigiendo errores tácticos, reconociendo los sutiles indicios de las posibilidades que se nos presentan, y controlando grandes proyectos que se prolongan durante meses o años. En los capítulos que siguen examinaremos con más detalle la arquitectura de esta máquina virtual, a fin de aportar nuevos apoyos —no pruebas— en favor de la hipótesis de que efectivamente puede asumir estas funciones ejecutivas y aun otras. Antes de hacerlo, no obstante, debemos poner al descubierto y neutralizar otra fuente de mistificación: la ilusión del Significador Central.

Una de las principales funciones de ese jefe imaginario es la de controlar la comunicación con el mundo exterior. Como vimos en el capítulo 4, la idealización que hace posible a la heterofenomenología *presupone* que hay alguien en casa para llevar el peso de la conversación, un autor de las grabaciones, un significador de todos los significados. Cuando nos disponemos a interpretar los sonidos vocales emitidos por un cuerpo locuaz, no suponemos que éstos son meros ladridos sin sentido, o palabras extraídas de un sombrero por una tropa de crápulas entre bastidores, sino que los tomamos como actos de un único agente, la (única) *persona* cuyo cuerpo está produciendo los sonidos. Si finalmente decidimos interpretar, no tenemos otra elección que postular la existencia de una persona cuyos actos comunicativos estamos interpretando. Esto no es exactamente lo mismo que postular un sistema *interno* que es el jefe del cuerpo, el titiritero que controla las marionetas, aunque sea ésta la imagen que naturalmente nos venga a la cabeza. Resulta tentador suponer que este jefe interno es un poco como el presidente de los Estados Unidos, que puede ordenar a su secretario de prensa o a cualquier otro subordinado que haga públicos los comunicados, pero cuan-

do éstos hablan, lo hacen en su nombre, ejecutan sus actos de habla, de los cuales es responsable, y de los cuales es, oficialmente, el autor.

De hecho, no hay ninguna cadena de mando como la descrita en el cerebro rigiendo la producción del habla (ni la de la escritura). Parte de la tarea de desmantelar el Teatro Cartesiano pasa por hallar una explicación más realista de cuál es la verdadera fuente de las aseveraciones, las preguntas y otros actos de habla que naturalmente atribuimos a esa (única) persona cuyo cuerpo está llevando a cabo las proferencias. Necesitamos ver qué le ocurre al mito capacitador de la heterofenomenología cuando se reconocen las complejidades de la producción del lenguaje.

Ya hemos podido ver alguna sombra cernirse sobre este problema. En el capítulo 4 imaginamos que Shakey el robot poseía una capacidad rudimentaria para conversar, o cuando menos de emitir palabras en determinadas circunstancias. Supusimos que Shakey podía diseñarse para «contarnos» cómo distinguía las cajas de las pirámides. Shakey podía decirnos «examino las secuencias de 10.000 dígitos...», o «busco fronteras de claro-oscuro y hago un dibujo lineal...», o «no lo sé; hay cosas que se me aparecen como cajas...». Cada uno de estos «testimonios» diferentes es el resultado de un diferente nivel de acceso que el dispositivo redactor de «testimonios» pudiera tener ante el funcionamiento interno del dispositivo identificador de cajas, pero no entramos en detalles sobre cómo los diferentes estados de la máquina podrían estar ligados a las diferentes producciones que éstos causaban. Aquél era un modelo deliberadamente simplificado de la producción del lenguaje, útil solamente para clarificar una cuestión muy abstracta a partir de un experimento mental: si un sistema emisor de enunciados tuviera sólo un acceso limitado a sus estados internos, y un vocabulario limitado con el cual componer esos enunciados, sus «testimonios» serían interpretables como verdaderos sólo si les imponemos una lectura metafórica. Las «imágenes» de Shakey constituían un ejemplo de cómo algo que no era realmente una imagen podía ser precisamente aquello de lo que se estaba hablando como si efectivamente se tratara de una imagen.

Una cosa es prever una nueva posibilidad abstracta; otra cosa es demostrar que esa posibilidad tiene una versión realista aplicable a nosotros mismos. Lo que Shakey hacía no era realmente un acto de emisión de testimonios, un acto de decir. Por lo que pudimos ver, esa verbalización imaginaria de Shakey podría ser perfectamente un amaño, ese lenguaje «enlatado» que los programadores incluyen en el *software* para hacerlo más atractivo al usuario. Usted va a formatear un disco y su ordenador le «hace» una pregunta amable: «¿Está usted seguro de querer hacer esto? Se borrará toda la información contenida en el disco. Responda S o N». Sólo un usuario muy ingenuo creería que el ordenador *pretendía* ser educado.

Permítanme poner algunas palabras en la boca de un crítico. Ya que este crítico imaginario seguirá los pasos de nuestras discusiones e investigaciones en capítulos posteriores, le pondremos un nombre. Habla Otto:

> Fue un truco muy barato el de referirse a Shakey cómo «él» en vez de «eso»; el problema con Shakey es que no tiene una interioridad verdadera como nosotros; no es posible imaginar cómo se sentiría uno si fuera eso. Aun en el caso de que los mecanismos responsables de recibir la información de entrada procedente de su «ojo» cámara de TV y de convertirla en identificación de cajas hubieran sido muy parecidos a los mecanismos de nuestro sistema visual (y no lo eran), y aun en el caso de que los mecanismos responsables de controlar la producción de cadenas de palabras inglesas hubieran sido muy parecidos a los mecanismos de *nuestro* sistema de producción del habla que controla la producción de cadenas de palabras inglesas (y no lo eran), *todavía* faltaría algo: ese intermediario en nuestro interior cuyos juicios se expresan cuando contamos cómo nos sentimos. El problema con Shakey es que sus entradas y salidas están conectadas de forma inadecuada, de una forma que elimina al observador (el experimentador, el beneficiario) que tiene que haber en algún punto entre la entrada visual y la salida verbal, de modo que haya alguien ahí dentro para *querer decir* las palabras de Shakey cuando son «proferidas».
>
> Cuando *yo hablo*, [prosigue Otto] quiero decir lo que digo. Mi vida consciente es privada, pero puedo decidir hacerle a usted partícipe de algunos de sus aspectos. Puedo decidir contarle cosas sobre mi experiencia pasada o presente. Cuando lo hago, formulo enunciados que adapto cuidadosamente al material que deseo relatar. Puedo ir de la experiencia al testimonio propuesto, contrastando las palabras con la experiencia para asegurar que he dado con *les mots justes*. ¿Tiene el sabor de este vino un matiz de *pomelo*, o quizá me recuerda más al de las *frambuesas*? ¿Sería mejor decir que ese sonido fuerte sonó *alto*, o quizá mejor diría que fue *claro* o *muy concentrado*? Atiendo a los dictados de mi experiencia consciente y llegó así a un juicio sobre las palabras que mejor le harán justicia a su carácter. Cuando me siento satisfecho de haber elaborado un testimonio preciso, lo expreso. A partir de mi relato introspectivo, usted puede llegar a conocer alguna característica de mi experiencia consciente.

Como heterofenomenólogos, hemos de dividir este texto en dos partes. Por una parte tenemos las afirmaciones que Otto hace sobre lo que experimenta cuando habla. Éstas son inviolables; así es cómo Otto percibe la experiencia, y debemos considerarlo como un dato susceptible de ser explicado. Por otra parte tenemos las afirmaciones teóricas (¿o acaso son conclusiones a argumentos tácitos?) que hace Otto sobre lo que todo esto demuestra que se está produciendo en su interior, y en qué medida es diferente de lo que está ocurriendo en el interior de Shakey, por ejemplo. Dichas afirmaciones no poseen ningún estatuto especial, pero las trataremos con el respeto que merece cualquier afirmación fruto de una reflexión.

Estoy totalmente de acuerdo en insistir en el hecho de que el intermediario, el observador interno en el Teatro Cartesiano, debe ser eliminado, de que nunca se hallará, pero no podemos *simplemente* deshacernos de él. Si no hay un Significador Central, ¿de dónde viene el significado? Debemos sustituirlo por una explicación plausible de cómo una proferencia que se quería decir —un testimonio real, sin comillas—, se compuso sin el *imprimátur* de un Significador Central solitario. Éste es el objetivo principal de este capítulo.

2. Burocracia frente a pandemónium

Uno de los cadáveres en el armario de la lingüística contemporánea es el haber sido muy prolija en todo lo referente al proceso de comprensión por parte de un oyente y el haber ignorado prácticamente el proceso de producción por parte de un hablante. Y sin embargo, podríamos afirmar que este proceso es la mitad del lenguaje, quizá la mitad más importante. Aunque existen muchas teorías detalladas y modelos de la *percepción* del lenguaje, y de la *comprensión* de enunciados percibidos por un oyente (el camino desde la fonología, pasando por la sintaxis, hacia la semántica y la pragmática), nadie —ni Noam Chomsky, ni ninguno de sus seguidores o de sus detractores— ha tenido nada particularmente interesante (aunque sea falso) que decir sobre los sistemas de *producción* del lenguaje. Es como si todas las teorías del arte fuesen teorías de la *apreciación* del arte, sin decir nunca una palabra sobre los artistas que lo crearon; como si todo el arte consistiera en *objets trouvés* apreciados por marchantes y coleccionistas.

No es difícil ver por qué esto es así. Los enunciados *son* objetos fáciles de encontrar para iniciar un proceso. Está realmente claro cuál es la materia prima o la entrada de los sistemas de percepción y comprensión: ciertos tipos de ondas en el aire, o grupos de marcas sobre una superficie plana. Y aunque hay una niebla bastante espesa que oscurece las controversias sobre cuál es el producto final del proceso de comprensión, al menos este profundo desacuerdo se produce al final del proceso que está siendo estudiado, no al principio. Una carrera con una línea de salida clara puede por lo menos empezarse, incluso si nadie está muy seguro de adónde llegará al final. ¿Es la «salida» o el «producto» de la comprensión del habla una *descodificación* o *traducción* de la entrada a una nueva representación —una oración en mentalés, quizá, o una imagen-en-la-cabeza— o es un conjunto de *estructuras profundas*, o una entidad que todavía nadie ha imaginado? Los lingüistas pueden decidir aplazar la respuesta a este rompecabezas mientras trabajan en las partes periféricas del proceso.

Con la producción del lenguaje, por otra parte, dado que nadie hasta el momento ha desarrollado una descripción clara y definitiva de lo que inicia el proceso que termina en una proferencia completa, es difícil incluso empezar a construir una teoría. Difícil, pero no imposible. Recientemente se han llevado a cabo algunos trabajos bastante interesantes sobre ciertos aspectos de la producción, sobre los cuales el psicolingüista holandés Pim Levelt ha realizado una excelente y muy bien organizada reseña en su libro *Speaking* (1989). Procediendo hacia atrás desde la salida, o procediendo desde la mitad en ambas direcciones, podemos echar algunas miradas bastante interesantes hacia los mecanismos responsables del diseño de nuestras proferencias y de que sean expresadas. (Los siguientes ejemplos se extraen del trabajo de Levelt.)

El habla no se produce en un «proceso por bloques» que diseña y ejecuta una palabra a la vez. La existencia de una capacidad de predicción en el

sistema, aunque sea limitada, se revela por la manera en que se distribuyen los acentos en el enunciado. Un caso simple: el acento de una palabra como «dieciséis» depende del contexto:

ANDY: ¿Cuántos dólares dices que cuesta?
BOB: Creo que dieciSÉIS.
ANDY: DIECiséis dólares no es mucho.

Cuando Andy profiere un segundo enunciado, debe ajustar su pronunciación de «dieciséis» a la palabra (DÓlares) que sigue. Si hubiera dicho:

dieciSÉIS no es mucho.

le habría dado a la palabra un patrón acentual distinto. Otro ejemplo: nótese la diferente acentuación en las dos apariciones de la palabra «Tennessee» en la frase

Conduje desde Nashville, TennessSEE, hasta la frontera de TENnessee y Arkansas.*

Los lapsus y otros errores lingüísticos demuestran de manera bastante concluyente cómo se respetan (y se dejan de respetar) las distinciones gramaticales en el momento de proyectar un enunciado que se va a emitir. Es más fácil oír a alguien decir «don los sueños» cuando quiere decir «son los dueños» que oírle decir «pin palabras» cuando quiere decir «sin palabras». Existe una mayor tendencia a utilizar las palabras reales (conocidas) frente a palabras meramente pronunciables (es decir, posibles pero no reales), incluso cuando se produce un *lapsus linguae*. Algunos errores son indicativos de cómo deben operar los mecanismos de selección de palabras: «Todo demonichucho [demonio+chucho] y diabliposa [diablo+mariposa]/en torno de su libro está volando»,* y «Acabo de poner el horno a velocidad baja». Piénsese también en el tipo de transposición implicada en la producción de errores como «¿Cuántos pesos quilas?» o «Si te acuerdo no me he visto», en vez de «¿Cuántos quilos pesas?» o «Si te he visto no me acuerdo».

Gracias a los ingeniosos experimentos que provocan tales errores, y a los complejos análisis de lo que ocurre y lo que no ocurre cuando la gente habla, se están realizando algunos progresos en la construcción de modelos

* Altero ligeramente el ejemplo para adecuarlo al orden de palabras castellano, ya que el fenómeno de corrimiento del acento (más claramente apreciable en inglés que en castellano) se produce cuando la palabra siguiente va acentuada en la primera sílaba, como ocurre con el nombre del estado de Arkansas (pronúnciese ARkansas). El ejemplo original debía traducirse «conduje desde Nashville, Tennessee, hasta la frontera de Tennessee» («I drove from Nashville, Tennessee, to the Tennessee border»), en cuyo caso «Tennessee» pasaba a ocupar la posición final y mantendría el acento agudo. [N. del T.]

* De unos versos de Quevedo. [N. del T.]

de los altamente organizados mecanismos que ejecutan la última articulación de un mensaje una vez *ya se ha decidido* que un determinado mensaje va a ser liberado al mundo exterior. Pero ¿qué o quién pone todos estos mecanismos en marcha? Un error de habla es un error precisamente en virtud de ser una cosa distinta de la que el hablante *quería decir.* ¿Qué capataz *asigna las tareas* en relación a las cuales se juzgan errores como los ejemplificados más arriba?

¿Qué, si no el Significador Central? Levelt nos propone un esquema, un «anteproyecto del hablante»:

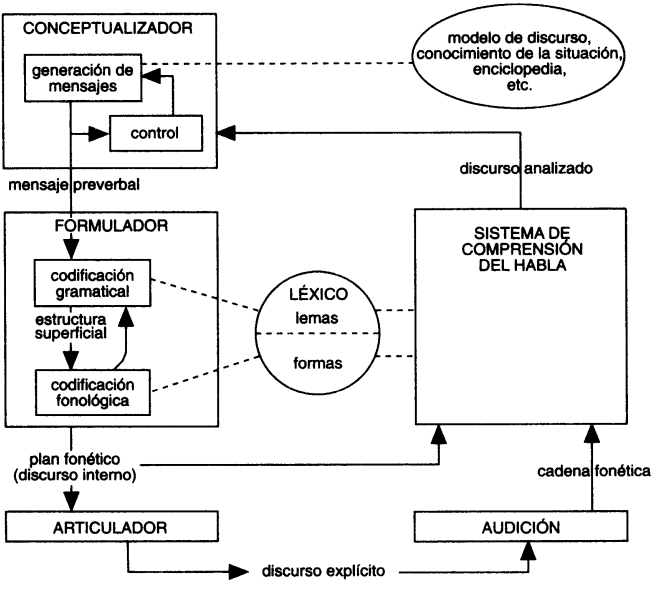


Figura 8.1

En el ángulo superior izquierdo hace su aparición un funcionario que tiene un sospechoso parecido con el Significador Central, armado con grandes cantidades de conocimiento del mundo, planes e *intenciones comunicativas,* y capacidad para la «generación de mensajes». Levelt advierte a sus lectores de que el conceptualizador «es una reificación que debe ser explicada» (pág. 9), pero postula su existencia de todos modos, ya que no parece poder hacer funcionar el proceso sin el recurso de un jefe, aun sin haberlo analizado, que dé las órdenes al resto del equipo.

¿Cómo funciona? El problema fundamental se verá más claramente si empezamos con una caricatura. El conceptualizador decide ejecutar un acto de habla, como por ejemplo insultar a su interlocutor haciendo un comen-

tario desagradable sobre sus pies. Envía, por tanto, la orden al grupo de burócratas bajo su influencia, el departamento de Relaciones Públicas (el formulador de Levelt): «¡Decidle a este burro que tiene los pies grandes!». Los chicos de RP se ponen al trabajo. Buscan las palabras apropiadas: la forma correcta en segunda persona del verbo *tener*, es decir *tienes*; un artículo masculino plural como *los*; una buena palabra para decir pies, como por ejemplo *pies*; y el adverbio y el adjetivo apropiados: *muy grandes*. Las combinan hábilmente con el tono de voz insultante adecuado y ejecutan la proferencia:

«¡Tienes los pies muy grandes!».

Pero esperen un momento, ¿es realmente tan sencillo? Cuando el conceptualizador dio la orden (lo que Levelt denomina *mensaje preverbal*), si la dio en castellano, como sugiere mi caricatura, entonces ya estaba hecho todo el trabajo difícil y apenas le quedaba nada por hacer al resto del equipo a parte de transmitir la orden con algún pequeño reajuste. Entonces, ¿acaso el mensaje preverbal está en un sistema de representación o lenguaje diferente? Sea lo que sea, debe ser capaz de proporcionar las «especificaciones» básicas al equipo de producción sobre el objeto que debe construir y liberar al exterior, y debe hacerlo en un formato que *este* equipo sea capaz de «comprender», quizá no castellano, pero sí alguna versión de cerebrés o mentalés. Debería ser una especie de lenguaje del pensamiento, argumenta Levelt, pero quizás en un lenguaje del pensamiento que se utilice únicamente para ordenar actos de habla, y no las demás actividades cognitivas. El equipo recibe el mensaje preverbal, una orden detallada en mentalés para producir un enunciado en castellano, y acto seguido cumple la orden. Así, los subordinados tienen algo más que hacer, pero con ello no hacemos más que ocultar la regresión que nos acecha. ¿Cómo decide el conceptualizador qué palabras en mentalés tiene que utilizar para dar la orden? Valdrá más que no tenga que haber un duplicado en miniatura del esquema de Levelt oculto en la cajita de *generación de mensajes* del conceptualizador (y así, sucesivamente, *ad infinitum*). Es obvio, además, que nadie le ha dicho al conceptualizador lo que tiene que decir; después de todo, él es el Significador Central, allí donde se *origina* el significado.

¿Cómo se desarrolla, pues, el significado de un enunciado? Considérese la siguiente serie de órdenes anidadas, partiendo de una estrategia general, pasando por tácticas más detalladas y hasta llegar a las operaciones básicas:

1. ¡Ponte ofensivo!
2. ¡Hazle algo feo, pero no demasiado peligroso!
3. ¡Insúltalo!
4. ¡Métete con alguna parte de su cuerpo!
5. ¡Dile que tiene los pies muy grandes!
6. Di: ¡tienes los pies muy grandes!
7. Profiere: *'tjenez los 'piez' 'muj' 'γrandes.*

No cabe duda de que debe pronunciarse una secuencia parecida a la descrita, en que todos los esfuerzos se dirigen hacia ese acto final. El habla humana es una actividad intencional; hay fines y medios, y llevamos a cabo un trabajo bastante aceptable en nuestro recorrido por las diversas opciones. Podríamos haberle dado un empujón en vez de insultarlo, podríamos haber despreciado su inteligencia en vez de su anatomía, o, citando a Fats Waller, podríamos haber dicho: «¡Tus extremidades pédicas son repulsivas!».

¿Pero es una jerarquía burocrática de jefes dando órdenes a sus subordinados la que lleva a cabo esta concentración de esfuerzos? En esta cascada de órdenes parece haber también mucha toma de decisiones, «momentos» en los que se «seleccionan» determinadas opciones en favor de otras opciones rivales, lo cual sugiere un modelo en el que se produce una delegación de responsabilidades para los más pequeños detalles, y en el que los agentes subordinados, con sus propias intenciones, evalúan los motivos para las diversas decisiones que toman. (Si no comprendieran por qué están haciendo lo que están haciendo, entonces no serían agentes, sino pasivos funcionarios de sello de goma que se dejan controlar por cualquier cosa que llega a su escritorio.)

El esquema de Levelt contiene vestigios de una de sus fuentes: la arquitectura de von Neumann que se inspiró en las reflexiones de Turing sobre su propio flujo de la conciencia y que, a su vez, ha inspirado muchos modelos en ciencia cognitiva. En el capítulo 7 intenté vencer la resistencia a la idea de que la conciencia humana es *muy parecida* a una máquina de von Neumann, un procesador serial con una sucesión de contenidos definidos pasando a través del cuello de botella del acumulador. Ahora debo echar un poco el freno y hacer hincapié en todo aquello en que la arquitectura funcional de la conciencia humana *no* se parece a la de una máquina de von Neumann. Si comparamos el esquema de Levelt con el modo en que normalmente una máquina de von Neumann emite palabras, vemos que el modelo de Levelt podría estar tomando prestadas demasiadas cosas.

Cuando una máquina de von Neumann dice lo que tiene escrito en su corazón, emite el contenido de su único espacio de trabajo central, el acumulador, el cual en cada momento siempre posee unos contenidos específicos en el lenguaje fijo de la aritmética binaria. Los rudimentarios «mensajes preverbales» de una máquina de von Neumann tienen una forma como la siguiente: 10110101 00010101 11101101. Una de las instrucciones primitivas de todo lenguaje máquina es una instrucción de SALIDA, la cual puede tomar el contenido actual del acumulador (por ejemplo, el número binario 01100001) y escribirlo en la pantalla o la impresora, de modo que un usuario exterior tenga acceso a los resultados obtenidos en la CPU. En una alternativa que facilita un poco más las cosas al usuario, una rutina compuesta por una serie de instrucciones primitivas traduce primero el número binario a notación decimal (por ejemplo, 00000110 en binario = 6 en decimal) o a una letra del alfabeto a través del código ASCII (por ejemplo, 01100001 en binario = «a» y 01000001 = «A») y después escribe el resultado. Subruti-

nas como éstas son las que encontramos en el centro de las más elaboradas instrucciones de salida de lenguajes de programación de alto nivel como el Fortran, el Pascal o el Lisp. Éstas permiten al programador crear nuevas subrutinas para construir mensajes más elaborados, yendo a buscar a la memoria largas series de números y haciéndolas pasar por el acumulador, traduciéndolas y escribiendo los resultados en la pantalla o la impresora. Por ejemplo, una subrutina puede efectuar diversos viajes al acumulador en busca de valores que le permitan rellenar los blancos en

> «Señor _____, tiene usted un descubierto en su cuenta de ptas. _____. Que tenga usted un buen día, señor _____».

Esta frase es una fórmula «enlatada» que a su vez se almacena en memoria como una serie de números binarios hasta que una subrutina determina que ha llegado el momento de abrir la lata. De este modo, una estricta jerarquía de rutinas fijas puede convertir secuencias de contenidos específicos dentro del acumulador en expresiones que un ser humano pueda leer en una pantalla o en una impresora: «¿Quiere usted guardar este documento?» o «6 archivos copiados» o «Hola, Billy, ¿quieres jugar al tres en raya?».

Hay dos propiedades de este proceso que el modelo de Levelt comparte: (1) el proceso toma como entrada un contenido predeterminado, y (2) la burocracia —el «flujo de control» en la jerga informática— debe estar diseñada con mucha precisión: toda la «toma de decisiones» fluye jerárquicamente por una delegación de responsabilidades a subagentes cuyo perfil determina qué parte del análisis de medios/fines están autorizados a llevar a cabo. Es interesante que la primera de estas propiedades —el contenido predeterminado— es algo que Otto parece ratificar al darnos la visión de sus propios procesos: en alguna parte, ahí en el centro, hay un «pensamiento» determinado que espera a ser «puesto en palabras». La segunda propiedad compartida, sin embargo, parece extraña: la jerarquía de rutinas que servilmente convierten *ese mismo* pensamiento en lenguaje natural han sido prediseñadas por *algún otro*: el programador, en el caso de la máquina de von Neumann y, presumiblemente, una combinación de evolución y desarrollo individual en el caso de las actividades del formulador de Levelt. El papel creativo y crítico que el pensador del pensamiento debería jugar al convertir éste en palabras no aparece en el modelo; o bien ha sido usurpado por el conceptualizador, que lleva a cabo todo el trabajo creativo antes de enviar una orden al formulador, o está implícito en el diseño del formulador, un *fait accompli* de un proceso previo de diseño.

¿De qué otra manera podrían organizarse los medios y los fines? Consideremos una caricatura opuesta: un pandemónium de palabras-demonio. Así es como hablamos: primero entramos en el modo de producir sonidos vocales, tocamos el claxon:

> «Piiii

No lo hacemos por ninguna razón en particular, sino simplemente porque no vemos ninguna buena razón para no hacerlo. Ese ruido «interno» excita a los demonios de nuestro interior que empiezan a intentar modular el ruido de manera aleatoria, interfiriendo en su flujo. El resultado es un galimatías, pero, al menos, un galimatías en castellano (en los hablantes del castellano):

> «Yaba-daba-duuuuu-duá-duá-tararí-tarará-pom-porom-pom...».

Pero antes de que todo este vocerío salga al exterior, otros demonios, sensibles a las regularidades en el caos, empiezan a darle forma de palabras, frases, muletillas...

> «Pues, ¿qué te parece?, el béisbol, ¿sabes?, de hecho, las fresas, por casualidad, ¿vale? Éste es el billete. Bueno, bien...».

Lo cual incita a nuestros demonios a realizar otros descubrimientos geniales, amplificados por un modelo oportunista, que producen secuencias más largas de verborrea un poco más aceptable, hasta que, finalmente, surge un enunciado:

«¡Te voy a dar una, que te vas a tragar todos los dientes!».

Afortunadamente, sin embargo, éste se descarta, no se profiere, ya que al mismo tiempo (en paralelo) se han ido urdiendo otros candidatos y ahora están por los alrededores, incluidos algunos claros perdedores, como por ejemplo

«¡Ay qué pillín!»

y

«¿Has leído algún libro últimamente?»

y un ganador por defecto, que acaba siendo proferido:

«¡Tienes los pies muy grandes!».

En esta ocasión, la musa le ha fallado a nuestro hablante; ninguna réplica chistosa llegó a la final, pero, cuando menos, acabó por surgir algo medianamente apropiado a la «disposición mental» actual del hablante. Cuando el hablante se vuelve a casa después del encuentro, probablemente retornará sobre ese torneo caótico, refunfuñando y meditando sobre lo que debería haber dicho. Puede que entonces la musa se presente con algo mejor, y el hablante lo saboreará, dándole vueltas y vueltas en su cabeza, imaginando

cómo se hubiera quedado su interlocutor si lo hubiera dicho. Al llegar a casa, puede que «recuerde» claramente haber espetado a su interlocutor una fresca bien afilada.

Podemos suponer que todo ello se produce en rápidas generaciones de procesamiento paralelo y «derrochador», con hordas de demonios anónimos y sus prometedoras construcciones que nunca verán la luz del día, sea porque son opciones que *conscientemente* son tomadas en consideración y rechazadas después, sea porque son actos de habla ejecutados finalmente para que los escuche un desconocido. Si hay tiempo suficiente, más de uno se probará en silencio en un ensayo consciente, aunque dichas audiciones formales son acontecimientos relativamente raros, reservados para aquellas ocasiones en las que hay mucho en juego y no conviene cometer errores. En una situación normal, el hablante no tiene pase previo; tanto él como su audiencia saben lo que va a decir en el mismo momento en que lo dice.

Pero, ¿cómo se arbitra este torneo de palabras? Cuando una palabra, una frase o una oración completa gana a sus competidores, ¿cómo se discriminan y valoran su conveniencia y su propiedad de acuerdo con la disposición mental actual? ¿Qué *es* una disposición mental (si no una intención comunicativa explícita) y cómo ejerce su influencia en el torneo? Ya que, después de todo, incluso si no hay ningún Significador Central, debe de haber alguna manera de que el contenido pase desde lo más profundo del sistema —desde los procesos perceptivos, por ejemplo— a las verbalizaciones.

Hagamos un repaso de estos asuntos. El problema con la hipótesis de la burocracia extrema es que el conceptualizador tiene demasiado poder, es un homúnculo con demasiados conocimientos y responsabilidades. Este exceso de poder se manifiesta en el difícil problema de cómo expresar su información de salida, el *mensaje preverbal*. Si ya especifica un acto de habla —si ya *es* una especie de acto de habla en mentalés, una orden específica para el formulador— lo más arduo de la difícil tarea de composición se habrá llevado a cabo antes de que nuestro modelo entre en acción. El problema con la alternativa del pandemónium es que necesitamos hallar un modo de que las fuentes del contenido puedan *influir* o *restringir* las energías creativas de los demonios-palabra sin *imponérselas*.

¿Qué hay del proceso descrito en el capítulo 1, los turnos de preguntas y respuestas que generaban alucinaciones en el modelo del juego del psicoanálisis? Recuérdese que eliminamos el sabio dramaturgo de sueños y productor de alucinaciones freudiano en favor de un proceso a partir del cual el contenido *surgía* bajo el incesante preguntar de un inquisidor. Nos quedaba el problema de cómo deshacernos del inteligente inquisidor, un problema que dejamos de lado. Ahora tenemos el problema complementario: cómo obtener respuestas para una muchedumbre de competidores haciendo preguntas como «¿Por qué no decimos: "¡tu madre lleva botas de militar!"?», o (en otro contexto) «¿Por qué no decimos: "me pareció ver un punto rojo moviéndose y volviéndose verde a medida que avanzaba"?». *Dos problemas complementarios... ¿será acaso posible que se resuelvan mutua-*

mente al tratarse juntos? ¿Y si los demonios-palabra fuesen, en paralelo, los inquisidores/competidores, y los demonios-contenido los contestadores/árbitros? Las intenciones comunicativas, completamente formadas y ejecutadas —los significados— podrían surgir de un proceso cuasievolutivo de diseño de actos de habla que comporta la colaboración, en parte serial, en parte paralela, de varios subsistemas, ninguno de los cuales es capaz por sí mismo de ejecutar —u ordenar— un acto de habla.

¿Es posible un proceso como éste? Hay toda una serie de modelos que contemplan dichos procesos de «satisfacción de condiciones», que ciertamente poseen propiedades sorprendentes. Además de la gran variedad de arquitecturas «conexionistas» con elementos parecidos a las neuronas (véase, por ejemplo, McClelland y Rumelhart, 1986), existen otros modelos más abstractos. La arquitectura Jumbo de Douglas Hofstadter (1983), que busca soluciones a jeroglíficos o anagramas, tiene las propiedades adecuadas, así como las tienen las ideas de Marvin Minsky (1985) sobre los agentes que conforman la «sociedad de la mente», que trataremos con más detalle en el capítulo 9. Debemos, sin embargo, abstenernos de emitir un juicio sobre el asunto, hasta que se hayan desarrollado y puesto en práctica modelos más detallados y explícitos, y claramente orientados hacia la producción del lenguaje. Puede haber sorpresas, buenas y malas.

Sabemos, no obstante, que en alguna parte de un modelo logrado de la producción del lenguaje deberemos valernos de un proceso evolutivo de generación de mensajes, ya que, de lo contrario, nos quedaremos atascados con un milagro («y entonces se produce un milagro») o con una regresión infinita de significadores que hagan el trabajo.[1] También sabemos —por las investigaciones que examina Levelt— que existen ciertos procesos automáticos y bastante rígidos que asumen el control al final, y que determinan las transformaciones de la estructura gramatical a la estructura fonológica, que componen la receta muscular final que termina en la palabra. Ambas caricaturas definen los extremos de un *continuum*, de lo hiperburocrático a lo hipercaótico. El modelo de Levelt —por contraste con la caricatura que he utilizado para hacer más clara la comparación— incorpora (o podría incorporar fácilmente) algunos rasgos no burocráticos de la caricatura opuesta: por ejemplo, ningún rasgo profundo o estructural impide al formulador de Levelt ocuparse en un proceso más o menos espontáneo (no solicitado, no

1. Dan Sperber y Deirdre Wilson (1986) abren una nueva perspectiva sobre cómo componemos nuestras comunicaciones, insistiendo en modelos de cómo *funcionan* realmente las cosas, en el hablante y en el oyente, contrariamente a la práctica habitual entre filósofos y lingüistas, que se limitan a hacer vagas alusiones a los mecanismos, al tiempo que apelan a reconstrucciones racionales de las supuestas tareas y sus requisitos. Ello permite a Sperber y Wilson hacer algunas consideraciones sobre la practicalidad y la eficiencia: principios del mínimo esfuerzo e interés por la organización temporal y la probabilidad. Demuestran, a continuación, cómo desde esta nueva perspectiva ciertos «problemas» tradicionales desaparecen, en particular el problema de cómo da el oyente con la interpretación «correcta» de las intenciones del hablante. Aunque no elevan su modelo al nivel de procesos evolutivos como los que hemos estado considerando aquí, sin duda invitan a llevar a cabo este desarrollo.

dirigido) de generación de lenguaje, y, dado el bucle de control entre el sistema de comprensión del habla y el conceptualizador (véase la figura 8.1), esta actividad espontánea *podría* jugar ese papel generador previsto para los demonios-palabra múltiples. Entre ambas caricaturas hay un todo, un abanico de maneras en que se podrían desarrollar modelos alternativos. La pregunta principal es ¿qué grado de interacción hay entre los especialistas que determinan el contenido y el estilo de lo que se debe decir y los especialistas que «conocen las palabras y la gramática»?

En uno de los extremos, la respuesta es: ninguna. Podríamos mantener intacto el modelo de Levelt y añadirle, simplemente, un modelo pandemónium de lo que ocurre *dentro* del conceptualizador para fijar el «mensaje preverbal». En el modelo de Levelt, hay una separación prácticamente total entre los procesos de generación de mensajes (elaboración de las especificaciones) y la producción lingüística (ejecución de esas especificaciones). Cuando el primer pedacito de mensaje preverbal llega al formulador, se desencadena la producción del principio de la proferencia y, a medida que el formulador va eligiendo las palabras, se van restringiendo las posibilidades de cómo puede continuar la proferencia, aunque existe un cierto grado de *colaboración* para la revisión de las especificaciones. Los ebanistas lingüísticos subordinados del formulador están, utilizando un término de Jerry Fodor, «encapsulados»; a su manera automática, hacen todo lo que pueden con las órdenes que reciben, sin un «si», ni un «y» ni un «pero».

En el otro extremo, tenemos modelos en los que las palabras y las frases del léxico, junto con sus sonidos, significados y asociaciones, avanzan a empellones con las construcciones gramaticales en un pandemónium, todas «intentando» ser parte del mensaje y haciendo así, algunas de ellas, una contribución sustancial a las intenciones comunicativas que sólo una pequeña minoría llegará a ejecutar. En este extremo, las intenciones comunicativas que existen son tanto un efecto del proceso como su causa; surgen como un producto y, una vez han surgido, quedan disponibles como patrones sobre los cuales se miden *ulteriores* implementaciones de las intenciones. No hay un único origen del significado, sino muchos orígenes cambiantes, desarrollados oportunamente a partir de la búsqueda de las palabras adecuadas. En lugar de un contenido predeterminado en un lugar funcional determinado, esperando a ser castellanizado por subrutinas, hay una disposición-mental-aún-incompletamente-determinada distribuida por el cerebro y restringiendo un proceso de composición que, con el tiempo, puede eventualmente realimentarse a fin de efectuar ajustes y revisiones, determinando aún más la misma tarea expresiva que puso en marcha el proceso de composición al principio. Siempre queda un patrón global de paso en serie, concentrado sobre un único tema a la vez, pero los límites son bastante difusos.

En el modelo en pandemónium, el control es usurpado en vez de ser delegado, en un proceso que en gran medida carece de diseño y es oportunista; hay múltiples fuentes para las «decisiones» de diseño que darán lugar a la proferencia final, sin que pueda existir una división del trabajo estricta

de las órdenes de contenido que fluyen desde el interior y las sugerencias voluntarias para la implementación hechas por los demonios-palabra. Lo que sugiere este modelo es que, a fin de *mantener* el papel creativo del *expresador*-de-pensamientos (algo que preocupaba mucho a Otto), debemos *abandonar* la idea de que el *pensador*-de-pensamientos empieza con un pensamiento determinado para ser expresado. A Otto también le importaba mucho esta idea de un contenido determinado, pero alguna cosa habrá que abandonar (y en la sección 4 examinaremos algunas alternativas con detalle).

¿En qué parte del espectro está la verdad? La respuesta debe ser determinada empíricamente y todavía no la conocemos.[2] Existen algunos fenómenos, sin embargo, que (a mi modo de ver) indican que la generación del lenguaje es un pandemónium —un proceso evolutivo, paralelo, oportunista— en casi todas sus etapas. En la próxima sección examinamos brevemente algunos de estos fenómenos.

3. Cuando las palabras quieren ser dichas

> No importa qué deseemos decir; es probable que no digamos exactamente *eso*.
>
> MARVIN MINSKY, *La sociedad de la mente*, 1985, pág. 236*

Los investigadores en inteligencia artificial Lawrence Birnbaum y Gregg Collins (1984) han observado una peculiaridad en los lapsus freudianos. Es sabido que Freud llamó la atención sobre los *lapsus linguae* que no son aleatorios ni carentes de sentido, sino, insistía Freud, profundamente significativos: inserciones inconscientemente *intencionadas* en el tejido del discurso, inserciones que satisfacen indirecta o parcialmente objetivos comunicativos suprimidos del hablante. La afirmación freudiana tradicional a menudo ha sido rechazada con gran vehemencia por parte de los escépticos, pero no hay nada sorprendente en su aplicación a casos particulares que no tienen nada que ver con la opinión que pueda tener uno sobre los temas más oscu-

2. Como señala Levelt, «Si se pudiera demostrar, por ejemplo, que la generación de mensajes se ve afectada directamente por la accesibilidad de los lemas o lexemas, dispondríamos de evidencias en favor de la idea de la realimentación desde el formulador al conceptualizador. Esta hipótesis puede ser evaluada empíricamente... Hasta ahora, la evidencia en favor de dicho proceso de realimentación ha sido negativa» (pág. 16). La evidencia que Levelt reseña proviene de experimentos muy controlados en los que se solicita al hablante que realice una tarea muy específica, como por ejemplo, *describir la imagen de la pantalla tan deprisa como pueda* (págs. 276-282). Como evidencia negativa es excelente —yo mismo me sorprendí ante la ausencia de efectos en estos experimentos—, pero, como el propio Levelt reconoce, es inconcluyente. No se puede sostener realmente que sea una afirmación *ad hoc* el decir que la artificialidad de estas situaciones experimentales ahogaba por completo la dimensión creativa/oportunista del lenguaje. Pero quizá Levelt esté en lo cierto; quizá la única realimentación entre el formulador y el conceptualizador sea *indirecta*: el tipo de realimentación que una persona puede producir *solamente* cuando se habla explícitamente a sí misma, formándose después una opinión sobre lo que se oye decir.

* Página 244 de la traducción castellana (Galápago, 1986). [N. del T.]

ros de la sexualidad que trata Freud, como el complejo de Edipo o las pulsiones de muerte. Freud comenta un ejemplo en el que un hombre dijo

«Señores, les invito a eructar a la salud de nuestro jefe».

(En alemán —la lengua en que originalmente fue proferido este enunciado— la palabra para decir «eructar» *aufzustoßen*, se coló en vez de la palabra para decir «brindar» *anzustoßen*.)

> En su explicación, Freud argumenta que este lapsus es la manifestación de un objetivo inconsciente por parte del hablante para ridiculizar o insultar a su superior, reprimido por las obligaciones políticas y sociales a las que debe hacer honor. Sin embargo, (...) no es razonable esperar que la intención del hablante de ridiculizar a su superior diera lugar en origen a un plan en el que se previera la palabra «eructar»: *a priori*, hay cientos de palabras y frases que podrían utilizarse con mayor plausibilidad para insultar o ridiculizar a alguien... No hay manera razonable de que el planificador hubiera podido anticipar que el objetivo de ridiculizar o insultar a su superior se vería satisfecho profiriendo la palabra «eructar», exactamente por la misma razón por la cual resulta improbable que el planificador hubiera escogido utilizar la palabra como insulto desde el principio.

El único proceso que podría explicar la frecuencia de esos felices y azarosos lapsus freudianos, argumentan Birnbaum y Collins, es el de una «planificación oportunista».

> ...Lo que parecen indicar ejemplos como el de más arriba, por tanto, es que los *propios* objetivos son agentes cognitivos, capaces de disponer de los recursos necesarios para reconocer las oportunidades de satisfacerse a sí mismos, y de los recursos comportamentales necesarios para sacar partido de esas oportunidades (Birnbaum y Collins, 1984, pág. 125).

Los lapsus freudianos destacan por el hecho de que parecen ser errores y a la vez no parecen serlo, aunque el hecho (si es que hay alguno) de que satisfagan objetivos *inconscientes* no los hace más difíciles de explicar que cualquier otro acto de elección de palabras que cumpla más de una función (u objetivo) al mismo tiempo. Es casi tan difícil imaginar cómo los juegos de palabras y otras formas intencionadas de humor verbal podrían ser el resultado de una planificación y una producción no oportunistas y encapsuladas. Si alguien tiene un plan para diseñar frases ocurrentes —un plan detallado que funcione de verdad—, hay más de un humorista que pagaría lo que fuera por poseerlo.[3]

3. Levelt me ha confesado ser un buscador empedernido de juegos de palabras (en holandés, su lengua materna) y, además, sabe cómo lo hace: «Gracias a muchos años de práctica, le doy vueltas prácticamente a cada palabra que oigo. Y después (de forma plenamente consciente), me fijo en el resultado para ver si significa algo. El 99,9 por ciento de las veces no surge nada que

Si Birnbaum y Collins están en lo cierto, el uso creativo del lenguaje sólo puede ser producto de un proceso paralelo en el que múltiples *objetivos* están simultáneamente al acecho de nuevos materiales. Pero, ¿y si los materiales mismos estuvieran al mismo tiempo al acecho de oportunidades para ser incorporados? Recogemos nuestro vocabulario de nuestra cultura; las palabras y las frases son nuestros rasgos fenotípicos más notables —los cuerpos visibles— de los memas que nos invaden, y no podría haber un medio más agradable en el que los memas puedan replicarse que un sistema de producción del lenguaje en el que los burócratas supervisores han dimitido en parte, cediendo una gran cantidad de control a las palabras mismas, quienes luchan entre ellas por una oportunidad de estar en el candelero de la expresión pública.

No es ninguna novedad que muchas de las cosas que decimos las decimos sobre todo porque nos gusta como suenan, no porque nos guste su significado. Nuevas formas de argot se extienden por las subcomunidades lingüísticas, abriéndose camino hasta el habla de casi todo el mundo, incluso de aquellos que intentan evitarlo. Muy pocos de los que utilizan una palabra nueva están siguiendo deliberada o conscientemente la máxima del maestro de escuela «Utiliza una nueva palabra tres veces y será tuya». A niveles superiores de agregación, incluso frases enteras nos atraen por cómo suenan a nuestros oídos o cómo vibran en nuestra lengua, independientemente de si cumplen alguna de las especificaciones proposicionales que habíamos decidido ejecutar. Una de las frases más citables de Abraham Lincoln es:

> Se puede engañar a todo el mundo alguna vez, y a algunas personas todas las veces, pero no se puede engañar a todo el mundo todas las veces.[4]

¿Qué quiso decir Lincoln? Los profesores de lógica gustan de señalar que en esta frase hay una «ambigüedad de alcance de los cuantificadores». ¿Quería Lincoln afirmar que hay algunos necios a los que siempre se puede engañar, o que en cada ocasión se puede engañar a una u otra persona, pero no siempre a la misma? Desde el punto de vista lógico, éstas son proposiciones totalmente distintas.

Compárese:

«Alguien siempre gana a la lotería».
«¡Pues debe estar forrado!»

---

sea divertido; pero en el uno por mil restante salen cosas bastante buenas y me falta tiempo para decirlas» (comunicación personal). Éste es un perfecto ejemplo de resolución de problemas von neumanniana: serial, controlada... *¡y consciente!* La pregunta es si hay otras maneras, más «pandemónicas», de generar gracias inconscientemente.

4. Según el *Oxford Dictionary of Quotations* (segunda edición, 1953), también se atribuye esta famosa frase a Phineas T. Barnum. Ya que Barnum es un ilustre alumno y generoso benefactor de mi universidad, me veo en la obligación de señalar que Lincoln podría no ser el originador de este mema de grandes capacidades replicativas.

«No es esto lo que yo quería decir.»

¿En cuál de las dos lecturas estaba pensando Lincoln? ¡Quizás en ninguna! ¿Qué probabilidades hay de que Lincoln nunca se percatara de la ambigüedad de alcance y de que nunca tuviera la ocasión de tener una intención comunicativa en vez de «la otra»? Quizá simplemente le sonó tan bien cuando la formuló por primera vez, que nunca percibió la ambigüedad, *y nunca tuvo ninguna intención comunicativa previa*, exceptuando la intención de decir algo sucinto y cargado de significado sobre el asunto de engañar a la gente. La gente habla así, incluso los grandes significadores como Lincoln.

La novelista Patricia Hampl, en un lúcido ensayo, «The Lax Habits of the Free Imagination», escribe sobre su propio proceso para componer relatos cortos.

> Toda historia tiene una historia. Esta historia secreta, con muy pocas oportunidades de ser contada, es la historia de su creación. Quizá la «historia de la historia» nunca pueda ser contada, ya que un trabajo acabado consume a su propia historia, la convierte en algo obsoleto, una vaina vacía (Hampl, 1989, pág. 37).

El trabajo acabado, observa Hampl, puede ser interpretado inmediatamente por los críticos como un artefacto hábilmente construido para satisfacer toda una serie de complejas intenciones del autor. Pero cuando ella se encuentra ante estas hipótesis sobre su propio trabajo, se siente azorada:

> «Hampl» tenía algunas intenciones preciosas, exceptuando, como la charlatana que de repente me sentí ser, la de sisar todo lo que hubiera encima de la mesa que fuera adecuado para mis propósitos. Aun peor, los «propósitos» eran vagos, inconsistentes, reversibles, estaban bajo presión. Pero, ¿quién —o qué— estaba haciendo esa presión? No podría decirlo (pág. 37).

¿Cómo lo hace, pues? Ella propone una máxima: «Seguid hablando; murmurar también sirve». Eventualmente, ese murmullo tomará formas que gocen de la aprobación del autor. ¿Es posible que el proceso que Hampl detecta a gran escala en su actividad creativa como escritora no sea más que una extensión del rápido y más profundo proceso que produce el habla creativa en la vida diaria?

Esta similitud tan tentadora no comporta solamente un proceso, sino también una actitud o una reacción subsiguiente. Ese celo que Hampl muestra en sus confesiones contrasta con la reacción más común —y no realmente deshonesta— de los autores a las interpretaciones cordiales de los lectores: estos autores muestran cierta deferencia ante las imputaciones de intencionalidad, e incluso se prestan a profundizar en ellas, con el espíritu que les puede hacer decir «¡Anda, creo que *eso* es precisamente lo que pretendía desde el principio!». ¿Y por qué no? ¿Acaso hay algo contradictorio en pensar que un movimiento que uno acaba de hacer (en el ajedrez, en la vida, en la

escritura) es finalmente mucho más inteligente de lo que pareció en un principio? (Para algunas reflexiones adicionales sobre este asunto, véase Eco, 1990a; 1990b.)

Como dijo E. M. Forster, «¿Cómo puedo decir lo que pienso, antes de ver lo que digo?». A veces sí que descubrimos lo que pensamos (y, por tanto, lo que queremos decir), reflexionando sobre lo que nos encontramos diciendo, y no corregimos. Así pues, al menos en estas ocasiones, estamos en el mismo barco que nuestros críticos e intérpretes exteriores, encontrándonos ante un pedacito de texto y asignándole la mejor lectura que podemos hallar. El hecho de que lo hayamos dicho le concede un cierto grado de capacidad de persuasión o, cuando menos, una presunción de autenticidad. *Probablemente* si lo dije (y me escuché decirlo, y no me oí apresurarme a corregirlo), es que quería decirlo, y seguramente significa lo que parece significar para mí.

La vida de Bertrand Russell nos ofrece un ejemplo:

> Ya era tarde cuando los dos invitados se despidieron y Russell se quedó solo con lady Ottoline. Permanecieron sentados junto al fuego hasta las cuatro de la madrugada. Russell, rememorando el acontecimiento unos días después, escribió, «no sabía que te amaba hasta que me oí decirlo; por un instante pensé "Dios mío, ¡qué he dicho!" e inmediatamente supe que era verdad» (Clark, 1975, pág. 176).

¿Qué decir de las demás ocasiones, no obstante, en las que no tenemos esa sensación de *descubrimiento*, de autointerpretación? Podríamos suponer que en estos casos, los normales, tenemos una íntima y privilegiada intuición previa de lo que queremos decir, por el simple motivo de que nosotros mismos somos significadores, la *fons et origo* del significado de las palabras que *nosotros* decimos, pero dicha suposición requiere un argumento que la sostenga, no basta con apelar a la tradición. Pues podría muy bien ser que no tuviéramos esa sensación de descubrimiento en estos casos simplemente porque para nosotros es tan obvio lo que queremos decir. No se necesita un «acceso privilegiado» para intuir que cuando digo, «Por favor, pásame la sal» durante la cena, estoy pidiendo la sal.

Hubo un tiempo en que no creía que hubiera alternativa al Significador Central, pero pensaba haber dado con un refugio para él. En *Content and Consciousness*, argumenté que debía de existir una línea divisoria funcionalmente clara (que denominé línea del conocimiento),* separando la fijación preconsciente de las intenciones comunicativas de su ejecución ulterior. La localización de dicha línea en el cerebro podía estar descaradamente amañada, anatómicamente hablando, pero tenía que existir, lógicamente, como la divisoria que separaba los funcionamientos defectuosos en dos va-

* Traduzco por «línea del conocimiento» la forma inglesa «awareness line», donde Dennett utiliza un sinónimo de «consciousness», a fin de mantener la diferencia de matiz que el autor busca. Utilizo «conocimiento» porque «awareness» en determinados contextos puede poseer este significado. [N. del T.]

riedades. Se podían producir errores en cualquier parte del sistema, pero todo error debía caer —por necesidad geométrica— a un lado u otro de la línea. Si caían del lado interior o superior de la línea, entonces *cambiaban aquello que iba a ser expresado* (el «mensaje preverbal» en el modelo de Levelt). El significado se fijaba en esta divisoria; de ahí es de donde provenía el significado. Debía existir un lugar como éste de donde proviniera el significado, pensaba yo, porque *algo* tiene que fijar los criterios con los cuales la «realimentación» pudiera registrar la incapacidad por ejecutar un acto de habla.

Mi error fue el de ser víctima de la misma ambigüedad de alcance que contamina la interpretación del dicho de Abe Lincoln. Es evidente que tiene que haber algo que en cada ocasión sea, por el momento, el criterio a partir del cual todo «error» corregido se corrija, pero no tiene por qué haber la misma y única cosa cada vez, incluso *dentro* de la duración de un acto de habla. No tiene por qué haber una línea *fija* (aunque esté amañada) que marque esa diferencia. De hecho, como vimos en el capítulo 5, la distinción entre revisiones preexperimentales que *cambian aquello que se experimentó* y revisiones postexperimentales que tienen el efecto de *relatar o registrar mal aquello que se experimentó* es indeterminada en extremo. A veces los sujetos se sienten inclinados a revisar o enmendar sus afirmaciones, y otras veces no. A veces, cuando efectúan revisiones, la narración corregida no está más cerca de «la verdad» o de «lo que *realmente* querían decir» que la versión anterior. Como señalamos anteriormente, el punto donde acaban las correcciones para la prepublicación y donde empieza la inserción de la fe de erratas para la postpublicación es una distinción que sólo puede establecerse de forma arbitraria. Cuando le hacemos una pregunta a un sujeto sobre si un determinado reconocimiento público captura o no captura adecuadamente la última verdad interior sobre lo que acaba de experimentar, el sujeto no se halla en mejor posición para juzgar de lo que lo están los observadores exteriores. (Véase también Dennett, 1990d.)

He aquí una nueva manera de observar el mismo fenómeno. Siempre que se produce el proceso de creación de una expresión verbal, al principio hay una distancia que debe ser salvada: lo que podríamos denominar la «distancia mal emparejada en el espacio semántico» entre el contenido que está en posición de ser expresado y los diversos candidatos para la expresión verbal que habían sido nominados. (En mi antigua visión, traté este problema como una mera cuestión de «corrección de realimentación», con un *punto fijo* para un rasero con el cual se iban a medir, descartar, mejorar los candidatos verbales.) Ese proceso de idas y venidas que reduce la distancia es una especie de proceso de realimentación, pero también es posible que el contenido-que-va-a-ser-expresado se ajuste en la dirección de alguna expresión candidata, ya que para cambiar o corregir la expresión candidata lo mejor es acomodar el contenido-que-va-a-ser-expresado. De este modo, las palabras y frases más accesibles o disponibles podrían realmente *cambiar el contenido de la experiencia* (si interpretamos la experiencia como lo que será rela-

tado en última instancia, el acontecimiento establecido en el mundo heterofenomenológico del sujeto).[5]

Si nuestra unidad como significadores no tiene más garantías que ésta, entonces en principio debería ser posible que esta unidad se quebrara en ciertas ocasiones. He aquí dos casos en los que esto es lo que parece haber ocurrido.

Una vez me vi en la desgraciada tesitura de tener que actuar como árbitro de primera base en un partido de béisbol; toda una novedad para mí. En el momento clave del juego (al final de la novena entrada, con dos fuera, y con la posibilidad de realizar la carrera final desde la tercera base), me tocó a mí decidir sobre la situación de un bateador que corrió hacia primera base. Era una jugada muy justa, y yo me encontré alzando enfáticamente el pulgar —la señal de FUERA— a la vez que gritaba «¡SALVADO!». Durante el tumulto que se organizó, todos me preguntaban qué había querido decir. Honestamente, no podía decirlo, al menos desde mi posición privilegiada. Finalmente, dado que no dominaba la señalización manual pero soy un hablante competente, decidí (para mí mismo) que era mi acto vocal el que debía prevalecer, aunque cualquiera podría haber realizado el mismo juicio. (Me gustaría saber de otras anécdotas en las que las personas no hayan sabido cuál de dos actos de habla diferentes querían ejecutar.)

Durante una situación experimental, el psicólogo Tony Marcel (en prensa) ha descubierto un ejemplo aún más elocuente. El sujeto, que padece de visión ciega (sobre la cual retornaremos en el capítulo 11), debía *decir* cuándo pensaba que se producía un destello de luz, pero las instrucciones que recibió sobre cómo debía hacerlo eran un tanto peculiares. Se le pidió que ejecutara ese único acto de habla mediante tres actos diferentes a la vez (no una secuencia, pero tampoco necesariamente «al unísono»):

1. decir «Sí»
2. presionar un botón (el botón del sí)
3. guiñar un ojo para decir sí

Lo sorprendente del caso es que el sujeto no siempre ejecutaba los tres actos juntos. Ocasionalmente guiñaba el ojo pero no decía SÍ o no presionaba el botón del SÍ, etc. No había una manera clara de ordenar las tres respuestas, bien según la fidelidad de la intención o según la precisión. Es decir, cuando había desacuerdo entre las tres acciones, el sujeto no tenía un claro criterio que le permitiera decidir qué acto debía aceptar y qué acto clasificar como un lapsus lingüístico, del dedo o del párpado.

5. Esto recuerda la visión de Freud sobre cómo funciona el «preconsciente»: «Así pues, la pregunta de cómo se hace algo consciente deberá ser sustituida por la de cómo se hace algo preconsciente, y la respuesta sería que por su enlace con las representaciones verbales correspondientes». (*El yo y el ello*, en la versión castellana de Luis López Ballesteros, Ramón Rey y Gustavo Dessal que aparece en el volumen *Los textos fundamentales del psicoanálisis*, Alianza, Madrid, 1988, pág. 555. [N. del T.])

Queda por ver si es posible provocar resultados similares bajo condiciones distintas y con otros sujetos, normales o no, aunque existen otras condiciones patológicas que parecen apuntar hacia un modelo de producción del lenguaje en el que la verbalización *puede* ponerse en marcha sin que haya órdenes enviadas por un Significador Central. Si usted padece alguna de estas patologías, «su mente está de vacaciones, pero su boca hace horas extraordinarias», como en la canción de Mose Allison.

La *afasia* es la pérdida o la merma de la capacidad de hablar. Existen numerosas variedades de afasias que son bastante comunes y han sido ampliamente estudiadas por neurólogos y lingüistas. En su variedad más común, la afasia de Broca, el paciente es plenamente consciente del problema y lucha, con un sentimiento creciente de frustración, por hallar las palabras que tiene en la punta de la lengua. En la afasia de Broca, la existencia de unas intenciones comunicativas frustradas es dolorosamente clara para el paciente. Sin embargo, en una variedad relativamente rara de afasia, la afasia de jerga, los pacientes no parecen mostrar ansiedad alguna sobre su déficit verbal.[6] Aunque su inteligencia es normal, y no padecen ningún tipo de psicosis o de demencia, se muestran totalmente satisfechos con proferencias verbales como las siguientes (extraídas de dos casos descritos en Kinsbourne y Warrington, 1963):

Caso 1:
¿Cómo estamos hoy?
«Chismorreando que muy bien y Lords y cricket y batallas de Inglaterra y Escocia. No lo sé. Hipertensión y dos ganaron al cricket, lanzar, batear y atrapar, pobrecillos, las cancelaciones pueden estar chismorreando, cancelaciones, brazo y argumento, acabando de lanzar.»
¿Cuál es el significado de «la seguridad ante todo»?
«Mirar y ver la Richmond Road en particular, y mirar el tráfico y dudar a la derecha y pasear, muy buena causa, quizás, las cebras pueden ser éstas, coches y el semáforo.»

Caso 2:
¿Trabajaba usted en una oficina?
«Sí, trabajaba en una oficina.»
¿Y qué tipo de empresa era?
«Oh, como ejecutivo, y la queja era discutir las tonaciones sobre de qué tipo eran, sobre cómo estaban impresas, y separadas de los diferentes... tricu... trículums, para ahorrarme los atributos conventementes... perdón...»

6. Levelt me informa de que recientes investigaciones que se están llevando a cabo en el Instituto Max Planck de Psicolingüística en Nimega ponen en duda esta afirmación, la tradicionalmente aceptada hasta ahora. Las investigaciones de Heeschen indican que, en un determinado nivel, los pacientes con afasia de jerga o de Wernicke sí sienten ansiedad por su déficit y parecen adoptar una estrategia de repetición con la esperanza de comunicarse.

«Ella quiere darle la vocación subjetiva para mantener la vocación de la perfecta impregnación de hermandad.»

«Su corrucación normal sería un punto.»

Se le ha pedido que identifique una lima de uñas:
«Esto es un cuchillo, una cola de cuchillo, un cuchillo, viejo, un viejo cuchillo.»*
y unas tijeras:
«Unas alamedas... son unas alamedas... no, no son unas alamedas... dos alamedas con un un peine... no, no es un peine... dos alamedas a condición de que el comandante no sea ahora...»

Una condición extrañamente similar, y mucho más común, es la *fabulación*. En el capítulo 4 propuse que las personas normales a veces pueden fabular sobre detalles de su propia experiencia, ya que son propensas a imaginar sin darse cuenta, y confunden el teorizar con el observar. La fabulación patológica es una ficción inconsciente de un tipo completamente distinto. Con cierta frecuencia, en algunos casos de lesión cerebral, especialmente cuando hay una terrible pérdida de memoria —como en el síndrome de Korsakoff (una secuela típica de casos muy graves de alcoholismo)— los pacientes no dejan de parlotear, contando mentiras flagrantes sobre sus vidas y sobre sus historias pasadas, e incluso sobre acontecimientos que se acaban de producir, si la amnesia es muy grave.

La verborrea resultante suena casi normal. A menudo suena igual que esa charla vacía y estereotipada propia de algunas conversaciones de bar: «Ah sí, mi mujer y yo... vivimos en la misma casa durante treinta años... solíamos ir a Coney Island, y, bueno, ya se sabe, nos sentábamos en la playa... nos *encantaba* sentarnos en la playa y mirar a los niños, y, bueno, pero eso era antes del accidente...». Sólo que todo es pura invención. La esposa puede haber muerto hace años, nunca estuvieron a menos de cien kilómetros de Coney Island, y pueden haber estado mudándose constantemente. Un oyente no iniciado a menudo no sabrá que se halla ante un fabulador, tan naturales y «sinceros» son sus recuerdos y tan ágiles son sus respuestas a las preguntas que se le hacen.

Los fabuladores no se dan cuenta de que lo están inventando todo, y los afásicos de jerga no son conscientes del hecho de que no hacen otra cosa que soltar una ensalada de palabras. Estas sorprendentes anomalías son ejemplos de *anosognosia*, o incapacidad por reconocer un déficit. Existen otras variedades de esta ausencia de autocontrol, y en el capítulo 11 consideraremos qué nos pueden decir sobre la arquitectura funcional de la con-

* En esta traducción literal se pierde casi todo el interés de esta proferencia, que se compone de una serie de palabras con una cierta similitud fonética con la palabra «nailfile» (lima de uñas). La proferencia original es la siguiente: «That is a knife, a knife tail, a knife, stale, stale knife».

ciencia. Por el momento nos limitaremos a señalar que la maquinaria cerebral es perfectamente capaz de construir actos de habla aparentes en ausencia de cualquier dirección coherente desde arriba.[7]

Las patologías, tanto si son pequeñas desviaciones inducidas por hábiles experimentadores como si son padecimientos más permanentes provocados por enfermedades o lesiones mecánicas del cerebro, nos proporcionan abundantes pistas sobre cómo están organizados los mecanismos. Dichos fenómenos me sugieren que nuestra segunda caricatura, el pandemónium, está mucho más cerca de la verdad de lo que lo estaría un modelo burocrático más dignificado, aunque todo esto aún tiene que ser verificado empíricamente. No estoy afirmando que para un modelo burocrático sea imposible hacer justicia a estas patologías, sino que éstas no parecen ser el tipo de fallos que uno esperaría en este tipo de sistemas. En el apéndice B, para científicos, mencionaré algunas líneas de investigación que pueden ayudar a confirmar o desmentir mis intuiciones.

En este capítulo he esbozado —pero no he probado— cómo un torrente de productos verbales, surgido a partir de miles de demonios-palabra formando coaliciones temporales, puede mostrar una cierta unidad, la unidad de una interpretación óptima en evolución, que hace que parezca *como si* este torrente fuera la ejecución de las intenciones de un conceptualizador; de hecho lo es, pero no de un conceptualizador *interno* que es parte integrante del sistema de producción del lenguaje, sino de un conceptualizador global, la persona, de la cual el sistema de producción del lenguaje es parte integrante.

Esta idea puede parecer extraña al principio, pero no debería sorprendernos. En biología hemos aprendido a resistir la tentación de explicar el *diseño de los organismos* apelando a la intervención de una única y gran inteligencia que lleva a cabo todo el trabajo. En psicología hemos aprendido a resistir la tentación de explicar la *visión*, diciendo simplemente que es como si hubiera un visionador de la pantalla interna, pues este visionador es el que hace todo el trabajo y lo único que se necesita entre este homúnculo y los ojos es una especie de cable de TV. También hemos ido construyendo una resistencia parecida a explicar la *acción* como algo que surge de los imperativos de un ordenador de acciones interno que lleva a cabo todo el trabajo de redactar las especificaciones. Como siempre, la manera de eliminar una inteligencia que es demasiado grande para nuestra teoría pasa por sustituirla, en última instancia, por un tejido mecánico de semiinteligencias semiindependientes actuando en conjunto.

Esta observación no sólo es aplicable a la generación de actos de habla;

7. Otro fenómeno lingüístico anómalo es ese conocido síntoma de la esquizofrenia: el «oír voces». Hoy en día es un hecho prácticamente establecido que la voz que «oye» el esquizofrénico es la suya propia; está hablando en silencio consigo mismo sin darse cuenta. Un obstáculo tan simple como hacer que el paciente se aguante la boca cerrada basta para que las voces callen (Bick y Kinsbourne, 1987). Véase también Hoffman (1986), y el comentario de Akins y Dennett, «Who may I say is calling?» (1986).

también es aplicable a toda acción intencional. (Véase Pears, 1984, donde se desarrollan ideas muy parecidas.) Contrariamente a lo que parecía en un principio, la fenomenología sí que nos ayuda a ver que es efectivamente así. Aunque en ocasiones somos conscientes de estar llevando a cabo un complejo razonamiento práctico que, una vez lo hemos tenido todo en cuenta, nos lleva a una conclusión sobre lo que debemos hacer, seguida de una decisión consciente de hacer eso, y que culmina finalmente con la ejecución del acto en cuestión, este tipo de experiencias son relativamente poco comunes. La mayoría de nuestras acciones intencionales se llevan a cabo sin preámbulos, lo cual es bueno, ya que, en caso contrario, no habría tiempo suficiente. La trampa en la que se suele caer consiste en suponer que esos raros casos de razonamiento práctico consciente son un buen modelo para el resto: esos casos en los que nuestras acciones intencionales surgen de procesos a los cuales no tenemos acceso. Por lo general nuestras acciones nos satisfacen; reconocemos que en gran medida son coherentes y que hacen contribuciones apropiadas y puntuales a nuestros proyectos, tal como los comprendemos. Así que, sin arriesgarnos, asumimos que éstas son el producto de procesos que, de manera bastante fiable, son sensibles a los fines y los medios. Es decir, son racionales, en *un* sentido de la palabra (Dennett, 1987a, 1991a). Pero ello no significa que sean racionales en un sentido más restringido: el producto de un razonamiento en serie. No tenemos que explicar los procesos subyacentes según el modelo de un agente interno que razona, concluye y decide, y que compara metódicamente los medios con los fines y entonces ordena una acción determinada; hemos visto un esbozo de cómo un tipo diferente de proceso podría controlar el habla y también nuestras acciones intencionales.

Sin prisa, pero sin pausa, vamos despojándonos de nuestros malos hábitos de pensamiento y sustituyéndolos por nuevos hábitos. La caída del Significador Central equivale a la caída del intencionador central, aunque el jefe sigue vivo bajo muy diferentes formas. En el capítulo 10 nos encontraremos con él en el papel de observador y reportero, y deberemos encontrar nuevas maneras de pensar en lo que está ocurriendo, pero antes debemos asegurar los fundamentos de nuestros nuevos hábitos de pensamiento, fijándolos más sólidamente a algunos detalles científicos.

Capítulo 9

LA ARQUITECTURA DE LA MENTE HUMANA

1. ¿Dónde estamos?

La parte más difícil ya ha pasado, pero todavía nos queda mucho trabajo por hacer. Hemos completado los más arduos ejercicios para ensanchar la imaginación, y estamos listos para poner a prueba nuestra recién estrenada perspectiva. Por el camino tuvimos que dejar muchos asuntos por tratar y nos permitimos una cierta desidia al tratar otros. Hay promesas que mantener, reconocimientos pospuestos y comparaciones que hacer. La teoría que he venido desarrollando incluye elementos tomados de muchos otros pensadores. En ocasiones, he ignorado deliberadamente aquello que estos pensadores consideran como la mejor parte de sus teorías, y he mezclado ideas procedentes de campos «hostiles», pero me he permitido omitir todos estos detalles en aras de una mayor claridad y llaneza. Es posible que hayamos dejado algunos serios modeladores de la mente retorciéndose de frustración, pero no se me ocurrió una manera mejor de hacer que diversos tipos de lectores lleguen al mismo punto de vista juntos. Ahora, sin embargo, estamos en una buena posición para hacer inventario y afianzar algunos detalles esenciales. Después de todo, la razón por la cual vale la pena tomarse la molestia de construir una nueva perspectiva es la de ver los fenómenos y las controversias de una manera diferente. Echemos, pues, una mirada.

En un pequeño esbozo, ésta es mi teoría hasta el momento:

> No hay un único y definitivo «flujo de la conciencia», porque no hay un cuartel general central ni un Teatro Cartesiano donde «todo se junta» para ser examinado por un Significador Central. En vez de este único flujo (por amplio que sea), hay múltiples canales en los que circuitos especializados intentan, en pandemóniums paralelos, llevar a cabo sus propias tareas, creando Versiones Múltiples a medida que avanzan. La mayor parte de estas versiones fragmentarias de «relato» juegan papeles efímeros en la modulación de la actividad del momento, aunque algunos se ven promocionados a nuevos papeles funcionales, en rápida sucesión, por la actividad de la máquina virtual en el cerebro. La serialidad de esta máquina (su carácter «von Neumanniano») no es un rasgo de diseño preconfigurado, sino el resultado de una sucesión de coaliciones entre estos especialistas.
>
> Los especialistas básicos forman parte de nuestra herencia animal. No se desarrollaron para llevar a cabo acciones propias de los humanos, tales como

> leer y escribir, sino para esquivar, evitar depredadores, reconocer caras, agarrar, lanzar, recoger bayas y otras tareas esenciales. A menudo se ven oportunamente alistados para nuevos papeles, para los cuales sus talentos originales son más o menos adecuados. El resultado no es un caos total simplemente porque las tendencias que se imponen sobre toda esta actividad son por sí mismas el producto de ese diseño. Parte de este diseño es innato y compartido con otros animales. Pero se ve ampliado, y a veces superado en importancia, por microhábitos de pensamiento que se han desarrollado en el individuo, en parte como resultado de la autoexploración individual y en parte como dones prediseñados de la cultura. Miles de memas, la mayor parte producida por el lenguaje, pero también «imágenes» sin palabras y otras estructuras de datos, fijan su residencia en un cerebro individual, conformando sus tendencias y convirtiéndolo, así, en una mente.

Esta teoría es lo bastante novedosa como para ser difícil de comprender, aunque se basa en modelos desarrollados en el seno de la psicología, la neurobiología, la inteligencia artificial, la antropología... y la filosofía. Este descarado eclecticismo a menudo provoca cierto recelo en los investigadores dentro de los campos de los que toma prestadas sus ideas. Gracias a mis frecuentes intrusiones en estas áreas, he llegado a acostumbrarme al poco respeto que sienten algunos individuos por sus colegas. «Pero Dan», me dicen los vinculados a la inteligencia artificial, «¿por qué pierdes el tiempo hablando con estos tipos de las neurociencias? Aluden al "procesamiento de la información" y sólo se preocupan por *dónde* se produce y sobre qué neurotransmisores intervienen, y todas estas cosas tan aburridas, pero no tienen ni la más remota idea de cuáles son los requisitos computacionales que imponen las funciones cognitivas superiores.» «¿Por qué —me preguntan los investigadores del cerebro— pierdes el tiempo con esas fantasías de la inteligencia artificial? Se dedican a inventar los mecanismos que les da la gana y demuestran una ignorancia imperdonable sobre todo lo que tiene que ver con el cerebro.» Sobre los psicólogos cognitivistas, por otra parte, recaen constantes acusaciones de construir modelos sin *ninguna* plausibilidad biológica *ni* poder computacional probado; los antropólogos no reconocerían un modelo si lo vieran, y los filósofos, como todos sabemos, se dedican a sacar los trapos sucios de unos y otros, advirtiendo sobre confusiones que ellos mismos han creado, en un área que carece por completo de datos y de teorías verificables empíricamente. Con tantos idiotas ocupándose del problema, no es de extrañar que la conciencia siga siendo un misterio.

Todas estas acusaciones son ciertas, y podríamos lanzar muchas otras todavía, aunque yo aún no me he cruzado con ningún idiota. Al contrario, la mayor parte de los teóricos cuyas ideas he tomado prestadas me parecen personas muy inteligentes —incluso brillantes, con la arrogancia y la impaciencia que a menudo acompaña a la brillantez—, pero con perspectivas y agendas limitadas. Son personas que intentan hacer progresos en el estudio de problemas muy complejos, tomando cualquier atajo que sean capaces de ver, al tiempo que deploran los que utilizan los demás. Nadie puede ser cla-

ro en el momento de tratar todos los problemas y todos los detalles, yo incluido, y todos nos vemos obligados a hablar entre dientes, a imaginar y a ser superficiales con grandes partes del problema.

Por ejemplo, una de las osadías endémicas en las ciencias del cerebro es la tendencia a pensar en la conciencia como si fuera *la parada terminal de la línea*. (Ello equivale a olvidar que el producto final de un manzano no son las manzanas, sino más manzanos.) Evidentemente, es sólo desde hace poco tiempo que estos investigadores se han permitido pensar en la conciencia, y sólo algunos excelentes teóricos han empezado a hablar, oficialmente, sobre lo que han pensado. Como ha comentado recientemente con cierta sorna el investigador de la visión Bela Julesz, uno sólo puede salir bien parado de todo este asunto si tiene el pelo cano... ¡y un Premio Nobel! Por ejemplo, he aquí una hipótesis aventurada por Francis Crick y Christof Koch:

> Hemos propuesto que una de las funciones de la conciencia sea la de presentar el resultado de diversas computaciones subyacentes y que ello comporta un mecanismo de atención que temporalmente enlaza las neuronas relevantes sincronizando sus picos en oscilaciones de 40 hz. (Crick y Koch, 1990, pág. 272).

Así que una de las funciones de la conciencia es *presentar los resultados de computaciones subyacentes*, pero ¿a quién? ¿A la reina? Crick y Koch no se plantean la pregunta difícil: *¿Y ahora qué?* («Y entonces, ¿se produce un milagro?») Una vez que su teoría ha conducido algo hacia lo que consideran como el círculo encantado de la conciencia, ésta se detiene. No afronta los problemas que tratamos en los capítulos del 5 al 8, por ejemplo, incluidos en particular los testimonios introspectivos.

Los modelos de la mente desarrollados en el seno de la psicología cognitiva y la IA, por el contrario, casi nunca tienen *este defecto* (véanse, por ejemplo, Shallice, 1972, 1978; Johnson-Laird, 1983, 1988; Newell, 1990). Generalmente postulan un «espacio de trabajo» o una «memoria de trabajo» que sustituye al Teatro Cartesiano, y muestran de qué manera los resultados de computaciones llevadas a cabo alimentan nuevas computaciones que rigen la conducta, informan los testimonios verbales, se doblan hacia atrás recursivamente a fin de proporcionar nueva información de entrada para la memoria de trabajo, etc. Sin embargo, estos modelos no suelen decir dónde o cómo podría localizarse una memoria de trabajo en el cerebro, y están tan ocupados con el *trabajo* que se lleva a cabo en ese espacio que no tienen tiempo para el «juego»; ningún signo de esa especie de *deleite* por la fenomenología que parece ser un rasgo tan importante de la conciencia humana.

Curiosamente, pues, los investigadores del cerebro a menudo se nos presentan como dualistas, ya que una vez han «presentado» las cosas a la conciencia, parecen cargarle el muerto a la mente, mientras que los psicólogos a menudo se nos presentan como zombistas (¿automatistas?), ya que describen estructuras que son desconocidas para los neuroanatomistas, y sus teo-

rías pretenden demostrar cómo se puede hacer todo el trabajo sin tener que recurrir a ningún observador interno.

Las apariencias engañan. Crick y Koch no son dualistas (aunque, aparentemente, son materialistas cartesianos), y los psicólogos cognitivistas nunca han negado la existencia de la conciencia (aunque la mayor parte del tiempo hacen todo lo posible por ignorarla). Además, ninguno de estos enfoques estrechos de miras descalifica a una u otra empresa. Los investigadores del cerebro tienen razón al insistir en que no se tiene un buen modelo de la conciencia hasta que no se ha resuelto el problema de cómo encaja este modelo en el cerebro, pero los investigadores en ciencia cognitiva (en inteligencia artificial y psicología cognitiva, por ejemplo) también tienen razón al insistir en que no se tiene un buen modelo de la conciencia hasta que se ha resuelto el problema de qué funciones ejecuta y de cómo las ejecuta, mecánicamente y no en beneficio de una mente. En palabras de Philip Johnson-Laird, «toda teoría científica de la mente debe tratar a ésta como si fuera un autómata» (Johnson-Laird, 1983, pág. 477). Las limitadas perspectivas de cada una de estas empresas no hacen más que demostrarnos la necesidad de una nueva empresa —aquella en que acabamos de embarcarnos— que intente aunar tantos puntos fuertes de las demás empresas como sea posible.

2. Orientándonos con el pequeño esbozo

Mi objetivo principal en este libro es de cariz filosófico: demostrar cómo *se podría* construir un modelo genuinamente explicativo de la conciencia con todos estos elementos diversos, y no desarrollar —y confirmar— dicha teoría en todos sus detalles. No obstante, mi teoría habría sido inconcebible (para mí, al menos), si no hubiera tomado prestados datos empíricos procedentes de diversos dominios, que abrieron el camino (para mí, al menos) a nuevas maneras de pensar. (Una compilación particularmente rica de nuevos descubrimientos e ideas sobre la conciencia es la de Marcel y Bisiach, 1988.) Vivimos unos tiempos apasionantes para la investigación de los problemas de la mente. El campo bulle con nuevos descubrimientos, nuevos modelos, sorprendentes resultados experimentales, y casi igual número de excesivamente aclamadas «pruebas» y refutaciones prematuras. En este momento, las fronteras de la investigación sobre la mente son tan amplias que apenas si podemos decir que haya un acuerdo sobre cuáles son las preguntas y los métodos apropiados. Con tantos fragmentos de teorías tan débilmente defendidos y tanta especulación, no es mala idea posponer nuestras exigencias de pruebas y demostraciones, y buscar por el contrario nuevos fundamentos, más o menos independientes, que tiendan a converger en apoyo de una única hipótesis, aunque sea de forma aún inconcluyente. Debemos intentar mantener a raya nuestro entusiasmo, sin embargo. A veces lo que parece bastante humo como para augurar un buen fuego, no es más que una nube de polvo provocada por el paso de una caravana.

En su libro *A Cognitive Theory of Consciousness* (1988), el psicólogo Bernard Baars establece lo que él considera como «el amplio consenso» de que la conciencia es el producto de una «sociedad distribuida de especialistas, equipada con una memoria de trabajo, denominada *espacio de trabajo global*, cuyo contenido es transmitido a todo el sistema» (pág. 42). Según observa Baars, una gran variedad de teóricos, a pesar de las grandes diferencias de perspectivas, formación y aspiraciones, gira en torno a esta visión compartida de cómo debe residir la conciencia en el cerebro. Lo que con cierta cautela he ido introduciendo aquí, no es más que una versión de esa visión compartida, ignorando algunas características y haciendo hincapié en otras, características que, a mi modo de ver, no han sido suficientemente tenidas en cuenta o han sido subestimadas, y que pienso que son cruciales para abrirnos paso hacia los misterios conceptuales que aún permanecen.

A fin de situar mi teoría en relación a algunos de los muchos trabajos de los cuales he tomado prestadas ideas, quisiera volver a mi pequeño esbozo, examinándolo tema por tema y estableciendo los paralelismos y los puntos de desacuerdo.

> No hay un único y definitivo «flujo de la conciencia», porque no hay un cuartel general central ni un Teatro Cartesiano donde «todo se junta» para ser examinado por un Significador Central...

Mientras casi todo el mundo coincide en rechazar la idea de que no hay un punto en el cerebro que se corresponda con la glándula pineal de Descartes, las implicaciones de dicha afirmación nunca han sido reconocidas y, ocasionalmente, se han pasado por alto de forma descarada. Por ejemplo, ciertas imprudentes formulaciones del «problema del ligamiento» en la investigación del cerebro actual a menudo presuponen que debe de haber un único espacio de representación en el cerebro (de menor tamaño que el propio cerebro), donde se aúnan las diferentes discriminaciones, casando la banda sonora con la película, coloreando las formas, rellenando las partes en blanco. Existen algunas formulaciones más cuidadosas del problema del ligamiento que evitan este error, pero a veces se pasan por alto los detalles.

> ...En vez de este único flujo (por amplio que sea), hay múltiples canales en los que circuitos especializados intentan, en pandemóniums paralelos, llevar a cabo sus propias tareas, creando Versiones Múltiples a medida que avanzan. La mayor parte de estas versiones fragmentarias de «relato» juegan papeles efímeros en la modulación de la actividad del momento...

Desde la IA, hace ya tiempo que Roger Schank viene señalando la importancia de las secuencias de tipo narrativo, primero en su trabajo sobre los *scripts* o *guiones* (1977, con Abelson), y más recientemente (1991) en su trabajo sobre el papel del relatar historias en la comprensión. Desde perspectivas muy distintas, también dentro de la IA, Patrick Hayes (1979), Marvin

Minsky (1975), John Anderson (1983) y Erik Sandeval (1991) —y otros—, han defendido la importancia de las estructuras de datos que no son meras secuencias de «instantáneas» (con el problema concomitante de reidentificar elementos particulares en los fotogramas sucesivos), sino que, de un modo u otro, están específicamente diseñadas para representar directamente secuencias temporales y tipos de secuencias. En la filosofía, Gareth Evans (1982), antes de su temprana muerte, había empezado a desarrollar ideas paralelas. En la neurobiología, estos fragmentos narrativos han sido estudiados, en tanto que *guiones* y otros tipos de secuencias, por William Calvin (1987) en su enfoque de la máquina darwinista. Hace tiempo que los antropólogos sostienen que los mitos que transmite cada cultura a sus nuevos miembros juegan un papel fundamental en la formación de sus mentes (véanse, por ejemplo, Goody, 1977, y, para algunas posibles aplicaciones a la IA, Dennett, 1991b), aunque nunca han intentado construir modelos computacionales o neuroanatómicos.

> ...aunque algunos se ven promocionados a nuevos papeles funcionales, en rápida sucesión, por la actividad de la máquina virtual en el cerebro. La serialidad de esta máquina (su carácter «von numanniano») no es un rasgo de diseño preconfigurado, sino el resultado de una sucesión de coaliciones entre estos especialistas...

Muchos han señalado el proceder lento y trabajoso de la actividad mental consciente (por ejemplo, Baars, 1988, pág. 120), y desde hace tiempo ha ido tomando cuerpo la idea de que ello se debe al hecho de que el cerebro nunca fue diseñado —preconfigurado— para tal actividad. Hace bastantes años, pues, que circula la idea de que la conciencia podría ser la actividad de una especie de máquina virtual serial implementada en el *hardware* paralelo del cerebro. El psicólogo Stephen Kosslyn presentó una versión de la idea de la máquina virtual en un congreso de la Society for Philosophy and Psychology a principios de los años ochenta, y yo mismo vengo intentando desarrollar diferentes versiones de esta idea desde hace aproximadamente el mismo tiempo (por ejemplo, Dennett, 1982b); sin embargo, una presentación aún más temprana de esta misma idea —aunque sin utilizar el término «máquina virtual»—, la encontramos en el artículo pionero del psicólogo Paul Rozin, «The evolution of intelligence and access to the cognitive unconscious» (1976). Otro psicólogo, Julian Jaynes, en sus audaces y originales especulaciones en *The Origins of Consciousness in the Breakdown of the Bicameral Mind* (1976), puso mucho énfasis en la idea de que la conciencia humana es una imposición muy reciente y producida por la cultura sobre una arquitectura funcional previa. El mismo tema, con algunas variaciones, también lo ha desarrollado el investigador del cerebro Harry Jerison (1973). De acuerdo con esta concepción, la arquitectura neuronal subyacente dista mucho de ser una *tabula rasa* o una pizarra en blanco en el momento del nacimiento, aunque es también un medio en el que se construyen estructuras, en fun-

ción de las interacciones del cerebro con el mundo. Y son estas estructuras que se construyen, más que las estructuras innatas, las que es preciso tomar en consideración para explicar el funcionamiento cognitivo.

> Los especialistas básicos forman parte de nuestra herencia animal. No se desarrollaron para llevar a cabo acciones propias de los humanos, tales como leer y escribir, sino para esquivar, evitar depredadores, reconocer caras, agarrar, lanzar, recoger bayas y otras tareas esenciales...

Muchas teorías diferentes coinciden en refrendar la presencia de estas hordas de especialistas, pero su tamaño, papel y organización son objeto de apasionados debates. (Para una útil y rápida reseña, véase Allport, 1989, págs. 643-647.) Los neuroanatónomos que estudian los cerebros de animales, desde las jibias y los calamares a los gatos y los monos, han identificado una gran variedad de circuitos preconfigurados, exquisitamente diseñados para llevar a cabo tareas específicas. Los biólogos hablan de Mecanismos Innatos de Desencadenamiento (Innate Releasing Mechanisms; IRM) y de Patrones de Acción Fijos (Fixed Action Patterns; FAP), que se pueden acoplar, y en una carta reciente, la neuropsicóloga Lynn Waterhouse describía muy acertadamente las mentes de los animales como «colchones de IRM-FAP». Son precisamente estas mentes animales problemáticamente acolchadas lo que Rozin (y otros) consideran como la base para la evolución de mentes con funciones más generales, que explotan estos mecanismos preexistentes para nuevos propósitos. El psicólogo de la percepción V. S. Ramachandran (1991), señala que «hay una ventaja real que aparece en los sistemas múltiples: proporciona tolerancia para las imágenes llenas de ruido como las que encontramos en el mundo real. Mi analogía favorita para ilustrar algunas de estas ideas es la de los dos borrachos: ninguno de los dos puede caminar por sí solo, pero apoyándose el uno en el otro consiguen alcanzar su objetivo a trancas y barrancas».

El neuropsicólogo Michael Gazzaniga ha aportado abundantes datos provenientes de ciertos déficits neurológicos (incluido el famoso, y a menudo mal descrito, caso de los pacientes con el cerebro dividido) que apoyan una concepción de la mente como una coalición o haz de acciones semiindependientes (Gazzaniga y Ledoux, 1978; Gazzaniga, 1985); y desde un campo diferente, el filósofo de la psicología Jerry Fodor (1983) ha argumentado que grandes partes de la mente humana se componen de *módulos*: sistemas preconfigurados, con una función específica y «encapsulados» para el análisis de la información de entrada (y la generación de información de salida, aunque Fodor apenas ha prestado atención a este punto).

Fodor concentra su atención sobre módulos que serían específicos de la mente humana —en particular, módulos para la adquisición del lenguaje y para analizar oraciones—, y, dado que en gran medida ignora el problema de cuáles podrían ser sus antepasados en las mentes de animales inferiores, crea, por así decir, la improbable impresión de que la evolución diseñó nue-

vos mecanismos propios para la especie. Esta imagen de dichos módulos como un milagroso don de la Madre Naturaleza al *Homo sapiens* se ve fomentada por la visión ultraintelectualista de Fodor sobre cómo los módulos están conectados al resto de la mente. Según Fodor, éstos no llevan a cabo tareas completas dentro de la economía de la mente (como por ejemplo controlar la coordinación de ojos y manos para recoger algo), sino que se detienen bruscamente en una frontera interna, una línea en la mente que no pueden cruzar. Fodor sostiene que existe un ruedo central para la «fijación racional de creencias», en la cual los módulos depositan servilmente sus bienes, convirtiéndolos así en procesos no modulares («globales, isotrópicos»).

Los módulos de Fodor son el sueño de un burócrata: la descripción de las tareas que deben llevar a cabo está grabada en piedra; no se puede acudir a ellos para que ejecuten tareas nuevas o jueguen más de un papel; y son «cognitivamente impenetrables», lo cual significa que sus actividades no pueden ser moduladas, ni interrumpidas, por cambios en los estados «informacionales» globales del resto del sistema. Para Fodor, todas las actividades que están realmente implicadas en el pensamiento *no* son modulares. Imaginar qué hacer después, razonar sobre situaciones hipotéticas, reestructurar los propios materiales de forma creativa, revisar la propia visión del mundo son, todas ellas, actividades ejecutadas por un misterioso dispositivo central. Además, Fodor afirma (mostrando un curioso sentimiento de satisfacción) que ninguna rama de la ciencia cognitiva, la filosofía incluida, tiene ni la más remota idea de cómo funciona este dispositivo central.

> Se sabe mucho sobre las transformaciones que se operan sobre las representaciones y que sirven para convertir la información a un formato que sea apropiado para el procesamiento central; apenas nada se sabe sobre qué le ocurre a la información una vez ha llegado hasta ahí. Hemos perseguido al espíritu hasta acorralarlo en lo más profundo de la máquina, pero todavía no lo hemos exorcizado. (Fodor, 1983, pág. 127.)

Al dotar a este dispositivo central con tantas cosas que hacer, y con tanto poder no modular con que hacerlo, Fodor convierte a sus módulos en unos agentes faltos de toda plausibilidad, agentes cuya existencia sólo tiene sentido en compañía de un agente jefe de una autoridad ominosa (Dennett, 1984b). Puesto que uno de los aspectos principales sobre los que Fodor hace hincapié al describir sus módulos es su mecanicidad finita comprensible y estúpida frente a los poderes inexplicables del centro no modular, aquellos teóricos que, en caso contrario, habrían sido receptivos a la mayor parte de su caracterización de los módulos han mostrado la tendencia a rechazarlos como las fantasías de un criptocartesiano.

Muchos de esos mismos teóricos han mostrado una actitud entre la indiferencia y la hostilidad para con los agentes de Marvin Minsky, quienes conforman la *Sociedad de la Mente* (1985). Los agentes de Minsky son homúnculos de muy diversos tamaños, desde gigantescos especialistas con talentos

tan sublimes como los de los módulos fodorianos, hasta agentes del tamaño de un mema (polinemas, micronemas,* agentes censores, agentes supresores, y muchos otros). Todo parece demasiado fácil, piensan los escépticos. Siempre que hay un trabajo por hacer, se postula un grupo de agentes ajustado a la tarea para llevarla a cabo; parafraseando a Bertrand Russell, es una estratagema teórica con todas las virtudes del latrocinio frente al trabajo honrado.

Los homúnculos —los demonios, los agentes— son moneda común en el reino de la inteligencia artificial y, en general, de las ciencias de la computación. Todo aquel que arruga la frente con aire escéptico ante la mención de los homúnculos simplemente no comprende hasta qué punto puede ser neutral este concepto, y cuán amplias sus aplicaciones. Postular un grupo de homúnculos sería efectivamente la estrategia fútil que el escéptico imagina, si no fuera por el hecho de que en las teorías homunculares, el verdadero contenido se halla en todo cuanto se dice sobre cómo los homúnculos interactúan, se desarrollan, forman coaliciones o jerarquías, etc. En cuanto a este punto, las teorías pueden ser muy distintas. Las teorías burocráticas, como vimos en el capítulo 8, organizan los homúnculos en jerarquías prediseñadas. No hay homúnculos subvencionados ni subversivos, y la competición entre homúnculos está tan regulada como una liga de béisbol. Las teorías del pandemónium, por el contrario, postulan la existencia de mucha duplicación de esfuerzos, derroche de movimientos, interferencias, períodos de caos y muchos gandules sin un trabajo definido. En estas teorías, llamar homúnculos (o demonios, o agentes) a estas unidades es casi tan poco significativo como llamarlas simplemente... unidades. Son sólo unidades con competencias bien definidas, y cada teoría, desde la que se cierne con más fidelidad a los datos neuroanatómicos hasta la más abstracta y artificial, postula algunas de estas unidades y después teoriza sobre el modo en que se pueden llevar a cabo funciones complejas a través de la organización de unidades que llevan a cabo funciones más simples. De hecho, todas las variantes del funcionalismo pueden verse como funcionalismos «homunculares» a un nivel de descripción de grano más o menos fino.

Me ha divertido observar una especie de eufemismo que últimamente ha ido calando entre los investigadores del cerebro. Los neuroanatónomos han hecho grandes progresos en la tarea de trazar un mapa del córtex, el cual resulta estar exquisitamente organizado en forma de columnas especializadas de neuronas que interactúan (que el investigador del cerebro Vernon Mountcastle, 1978, denomina «módulos unidad»), organizadas a su vez en organizaciones mayores tales como los «mapas retinotópicos» (en los que se conserva el patrón espacial de excitación sobre las retinas de los ojos),

* No confundir los memas con los nemas; éstos últimos los define Minsky, en *La sociedad de la mente*, como «agentes cuyas salidas afectan a otros agentes en maneras que cada uno aprende independientemente. El "contexto" dentro del cual trabaja un agente típico está determinado en gran parte por la actividad de los nemas que llegan a él». (Minsky, 1985, trad. cast., pág. 346; N. del T.)

que a su vez juegan un papel —apenas conocido aún— en organizaciones aún mayores de neuronas. Los investigadores del cerebro solían hablar de lo que estos distintos trazados o grupos de neuronas *señalaban*; pensaban en estos grupos como si fueran homúnculos cuyo «trabajo» era siempre «enviar un mensaje con un contenido determinado». Nuevas y recientes reflexiones sobre el asunto apuntan hacia la idea de que en dichas regiones se llevan a cabo funciones mucho más complejas y variadas, de modo que hoy en día se considera un gran error el hablar de ellas (únicamente) como meras señaladoras de esto o aquello. ¿Cómo podríamos expresar, entonces, esos descubrimientos, que tanto trabajo han costado, sobre las condiciones específicas bajo las cuales se activan dichas regiones? Decimos que esta región «se interesa» por el color, mientras que aquella «se interesa» por la localización o el movimiento. Pero este uso no es el ridículo antropomorfismo o «la falacia del homúnculo» que nos encontramos por todas partes en la IA, por supuesto; no es más que una manera inteligente e imaginativa de hablar sobre las funciones de las regiones nerviosas, ideada por unos científicos serios. Lo que es bueno para lo uno es bueno para lo otro.

Los agentes de Minsky son distintos principalmente porque, a diferencia de otras variedades de homúnculos, tienen una historia y una genealogía. Su existencia no sólo se postula, sino que se pueden haber desarrollado a partir de algo previo cuya existencia no constituía un misterio; Minsky tiene muchas sugerencias sobre los posibles desarrollos que pueden haberse producido. Si sigue manteniendo una actitud desconcertantemente evasiva sobre qué tipos de neuronas podrían ser los constituyentes de sus agentes y sobre cuál sería su localización en el cerebro, es porque ha querido estudiar los requisitos más generales sobre el desarrollo de las funciones, sin caer en el exceso de detalle. Como él mismo señala, al describir su anterior teoría de los *frames* (cuyo descendiente es la sociedad de la mente), «Si la teoría hubiera sido algo más vaga, habría sido ignorada, pero si hubiera sido descrita con mayor detalle, tal vez otros científicos la habrían puesto [a prueba], en lugar de aportar su propias ideas» (1985, pág. 259).* Algunos científicos no se sienten particularmente afectados por este tipo de apologías. Sólo se interesan por aquellas teorías capaces de hacer predicciones verificables de inmediato. Ésta sería una buena política, si no fuera por el hecho de que todas las teorías verificables desarrolladas hasta ahora son manifiestamente falsas, y sería absurdo pensar que los avances necesarios para construir *nuevas* teorías verificables surgirán de la nada, sin una buena dosis de reflexión imaginativa como la que se permite Minsky. (Evidentemente, yo estoy jugando al mismo juego.)

Volvamos a nuestro pequeño esbozo:

> A menudo [los demonios especialistas] se ven oportunamente alistados para nuevos papeles, para los cuales sus talentos originales son más o menos ade-

* Página 267 de la traducción castellana (Galápago, 1986). [N. del T.]

cuados. El resultado no es un caos total simplemente porque las tendencias que se imponen sobre toda esta actividad son por sí mismas el producto de ese diseño. Parte de este diseño es innato y compartido con otros animales. Pero se ve ampliado, y a veces superado en importancia, por microhábitos de pensamiento que se han desarrollado en el individuo, en parte como resultado de la autoexploración individual y en parte como dones prediseñados de la cultura. Miles de memas, la mayor parte producida por el lenguaje, pero también «imágenes» sin palabras y otras estructuras de datos, fijan su residencia en un cerebro individual, conformando sus tendencias y convirtiéndolo, así, en una mente.

Es en esta parte de mi teoría donde me he mostrado deliberadamente evasivo sobre muchas cuestiones importantes: ¿de qué manera interactúan estos homúnculos a fin de llevar algo a cabo? ¿Cuáles son las transacciones de procesamiento de la información subyacentes, y qué motivos tenemos para pensar que podrían «funcionar»? De acuerdo con mi esbozo, la secuencia de acontecimientos se ve determinada (de maneras a las que sólo he hecho alguna alusión indirecta) por «hábitos», y, aparte de mis aseveraciones negativas en el capítulo 5 sobre lo que *no* ocurre, apenas he hecho afirmaciones específicas sobre la estructura de los procesos a través de los cuales se perpetúan elementos fruto de las Versiones Múltiples, algunos de los cuales acabarán por generar heterofenomenología como resultado de un sondeo u otro. A fin de poder comprobar adónde nos lleva todo esto, y cuáles podrían ser las diferentes respuestas alternativas, deberíamos echar una rápida mirada a algunos modelos más del pensamiento secuencial.

3. ¿Y AHORA QUÉ?

En el capítulo 7 pudimos comprobar que la arquitectura de von Neumann es el producto de un proceso serial de cálculo deliberado. Turing y von Neumann aislaron un tipo particular de corriente que puede fluir a través del flujo de la conciencia y, después, lo idealizaron de forma radical en interés de la mecanización. Tenemos el célebre cuello de botella de von Neumann, que consiste en un único registro para los resultados y otro para las instrucciones. Los programas no son más que listas de instrucciones ordenadas, construidos a partir de un pequeño conjunto de primitivos que la máquina está preconfigurada para ejecutar. Un proceso fijo, el ciclo de buscar y ejecutar, extrae las instrucciones de una lista en la memoria, una por vez, tomando siempre la primera de la lista, a menos que la instrucción precedente no provoque un desvío hacia otra parte de la lista.

Cuando los investigadores en IA comenzaron a construir modelos más realistas de las operaciones cognitivas sobre esta base, llevaron a cabo una importante revisión de la misma. Expandieron el escandalosamente estrecho cuello de botella de von Neumann y lo convirtieron en un «espacio de trabajo» o una «memoria de trabajo» más compendioso. También diseñaron

operaciones más sofisticadas que funcionaran como primitivas psicológicas, y sustituyeron el rígido ciclo de buscar y ejecutar de la máquina de von Neumann por maneras más flexibles de llamar a las instrucciones para ser ejecutadas. El espacio de trabajo se convirtió, en algunos casos, en una «pizarra» (Reddy y otros, 1973; Hayes-Roth, 1985), sobre la que los diversos demonios pudieran dejar mensajes para ser leídos por los demás demonios, lo cual a su vez provocaba un nuevo turno de escritura y lectura de mensajes. La arquitectura de von Neumann, con su rígido ciclo de instrucciones, seguía ahí, en un último término, llevando a cabo la implementación, pero no jugaba ningún papel en el modelo. En el modelo, lo que ocurría después estaba regido por los resultados de oleadas competitivas de escritura y lectura de mensajes sobre la pizarra. Otra especie vecina de descendientes de la arquitectura de von Neumann es la formada por los diversos *sistemas de producciones* (Newell, 1973) que subyacen a modelos tales como el ACT* (léase «act-star») de John Anderson (1983) y el Soar de Rosenbloom, Laird y Newell (1987); véase también Newell, 1990.

Pueden ustedes hacerse una idea bastante clara de cuál es la arquitectura subyacente a un sistema de producciones a partir de este esquema simple de ACT* (figura 9.1).

La memoria de trabajo es el lugar donde se llevan a cabo todas las acciones. Todas las acciones básicas se denominan *producciones*. Las producciones son, en esencia, mecanismos de reconocimiento de patrones ajustados para dispararse cada vez que detectan *su* patrón. Es decir, existe una serie de operadores (SI-ENTONCES) que circulan examinando el contenido de la memoria de trabajo, a la espera de que sus condiciones se vean satisfechas (SI), a fin de poder (ENTONCES) llevar a cabo sus actos, sean cuales sean (en un sistema de producciones clásico, suele tratarse de depositar un nuevo elemento en la memoria de trabajo, para el examen ulterior de otras producciones).

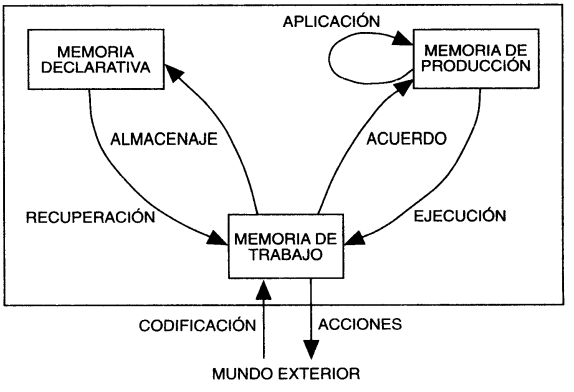


Figura 9.1

Todos los ordenadores poseen primitivas (SI-ENTONCES), los «órganos sensoriales» que les permiten reaccionar de forma diferenciada a datos procedentes o recuperados de la memoria. Esta capacidad para la *ramificación condicional* es un ingrediente esencial del poder de un ordenador, independientemente de cuál sea su arquitectura. En su origen, los condicionales eran las claras y simples instrucciones de transición de estado de una máquina de Turing: SI lees un cero, ENTONCES sustitúyelo por un uno, desplázate un espacio hacia la izquierda, y pasa al estado *n*. Compárense estas instrucciones tan simples con las que se le podrían dar a un centinela humano bien entrenado y experimentado: SI ves algo que no te resulta familiar, y un examen más detallado no resuelve la cuestión O te queda alguna duda, ENTONCES haz sonar la alarma. ¿Podemos construir un sistema de control tan elaborado como éste a partir de simples y mecánicos condicionales? Las *producciones* son sensores de nivel intermedio a partir de los cuales se podrían construir órganos sensoriales más complejos y, a partir de ahí, arquitecturas completas de la cognición. Las producciones pueden tener prótasis complejas y difusas; los patrones que «reconozcan» no tienen que ser tan simples como los códigos de barras que leen las cajas registradoras, sino más bien como los patrones que podría discriminar un centinela (véanse los comentarios al respecto en Anderson, 1983, págs. 35-44). Y a diferencia de los SI-ENTONCES de una máquina de Turing, que siempre se encuentra en un estado diferente cada vez (y siempre comprobando uno de los condicionales del conjunto, antes de pasar a examinar el siguiente dato), los SI-ENTONCES de un sistema de producción esperan *en masse*, en paralelo (simulado), de modo que, en cada «instante», más de una producción puede ver satisfecha su condición y estar lista para actuar.

Aquí es donde las cosas se ponen interesantes: ¿cómo tratan este tipo de sistemas la *resolución de conflictos*? Cuando más de una producción se ve satisfecha, siempre existe la posibilidad de que dos (o más) empujen en direcciones contrarias. Los sistemas paralelos pueden tolerar una gran cantidad de objetivos cruzados, pero en un sistema que debe tener éxito en el mundo, no todo puede ocurrir al mismo tiempo; a veces, algo tiene que ceder. La manera de tratar la resolución de conflictos es un aspecto clave en el que los modelos difieren. De hecho, ya que la mayoria, si no todos, de los detalles psicológica y biológicamente interesantes tienen que ver con diferencias a este nivel, resulta conveniente considerar la arquitectura de los sistemas de producciones como el medio a partir del cual construir modelos. Todos los sistemas de producciones poseen, no obstante, un rasgo en común que nos permite establecer un puente con nuestro esbozo de teoría: todos poseen un espacio de trabajo donde se llevan a cabo las acciones, donde muchas producciones (= demonios) pueden intentar hacer su trabajo a la vez, y poseen, además, una memoria más o menos inerte donde se almacena la información innata y la información que se ha ido acumulando. Dado que no todo lo que «sabe» el sistema está disponible en el espacio de trabajo a la vez, el problema de Platón de hacer que acuda el pájaro adecuado en el

momento apropiado se convierte en el principal problema logístico al que hay que enfrentarse. Y lo que es aún más importante, los teóricos *han desarrollado efectivamente* mecanismos con el objetivo de responder a la pregunta difícil: *¿y ahora qué?*

Por ejemplo, en ACT* existen cinco principios básicos de resolución de conflictos:

1. *Grado de acuerdo*: si la prótasis de una producción encaja mejor sobre el patrón que las de otras, entonces éste tiene prioridad sobre ellas.
2. *Fuerza de producción*: las producciones que han tenido éxito recientemente tienen una mayor «fuerza» asociada, lo cual les concede prioridad sobre producciones con fuerza menor.
3. *Refrangibilidad a los datos*: una misma producción no puede encajar con los mismos datos más de una vez (esto es para prevenir bucles infinitos y, en casos menos drásticos, rutinas similares).
4. *Especificidad*: cuando dos producciones encajan con los mismos datos, la producción con una prótasis más específica tiene prioridad.
5. *Dominio de los objetivos*: entre los elementos que las producciones depositan en la memoria de trabajo hay objetivos. Sólo puede haber un objetivo activo por vez en la memoria de trabajo de ACT*, y toda producción cuya salida encaje con el objetivo tiene prioridad.

Todos estos principios resultan plausibles para la resolución de conflictos, pues tienen sentido tanto desde el punto de vista psicológico como desde el punto de vista teleológico (para un discusión detallada de este punto, véase Anderson, 1983, cap. 4). Pero quizá tengan demasiado sentido. Es decir, el mismo Anderson diseñó inteligentemente el sistema para la resolución de conflictos de ACT*, explotando sus conocimientos sobre los tipos específicos de problemas que surgen en estas circunstancias, y sobre métodos efectivos para tratarlos. En esencia, ha preconfigurado este complejo sistema de conocimiento, un don innato de la evolución. Como contraste, es interesante considerar la arquitectura del sistema Soar de Rosenbloom, Laird y Newell (1987). Éste, como cualquier sistema de arquitectura paralela, se topa con *impasses*, ocasiones en las que se necesita resolver conflictos sea porque «se disparan» producciones contradictorias o no «se dispara» ninguna, pero los trata como ventajas, no como problemas. Estos callejones sin salida son esenciales para construir oportunidades para el sistema. Los conflictos no se abordan automáticamente mediante un conjunto fijo y clarividente de principios de resolución de conflictos (un autoritario homúnculo con funciones de guardia urbano que está ahí desde el principio), sino que se tratan de manera *no* automática. Un callejón sin salida crea un *nuevo* «espacio de problemas» (como un espacio de trabajo para los casos de mayor actualidad) en el que el problema que hay que resolver es precisamente el *impasse* con que se ha topado el sistema. Esto puede, a su vez, generar otro metametaespacio de problemas de tráfico y así sucesivamente, *potencialmen-*

te para siempre. Pero, en la práctica (cuando menos en los dominios modelados hasta el momento), después de haber apilado bastantes espacios de problemas el uno sobre el otro el problema que está más arriba se ve resuelto, lo cual sirve rápidamente para resolver el siguiente problema en la pila, y así sucesivamente, eliminando esa terrible proliferación de espacios después de haber llevado a cabo un examen nada trivial del espacio lógico de posibilidades. Asimismo, los efectos sobre el sistema consisten en «trocear» los descubrimientos hechos a duras penas resultantes en nuevas producciones, de modo que cuando en el futuro surjan problemas similares, ya se ha acuñado una nueva producción que servirá para resolver rápidamente un problema que ya se había resuelto con anterioridad.

Al mencionar brevemente estos detalles, no es mi intención argumentar en favor de Soar y en contra de ACT*, sino dar una idea del tipo de cuestiones que se pueden explorar, de modo responsable, mediante modelos construidos con este tipo de partes. Intuyo, por razones que no nos interesan ahora, que el medio subyacente a los sistemas de producciones *todavía* está demasiado idealizado y es excesivamente simple en cuanto a sus restricciones se refiere. Sin embargo, el camino desde la máquina de von Neumann hasta los sistemas de producciones apunta hacia nuevos tipos de arquitecturas, aún más parecidas a la del cerebro, y la mejor manera de estudiar sus posibilidades y sus limitaciones es construirlas y hacerlas funcionar. Ésta es la manera de convertir lo que, en teorías como la mía, todavía es impresionista y vago en modelos honestos que se puedan verificar empíricamente.

Cuando se toman las diferentes afirmaciones sobre los mecanismos de la conciencia que he venido defendiendo en los últimos cuatro capítulos y se empieza a intentar yuxtaponerlas sobre modelos de sistemas cognitivos como éstos, surgen numerosas preguntas que no voy a intentar responder aquí. Dado que dejo todas las preguntas por resolver, mi esbozo se queda en eso, en un esbozo que casaría con toda una familia de teorías diferentes. En esta ocasión, me basta con esto para seguir adelante, ya que los problemas filosóficos de la conciencia consisten en saber si *alguna* teoría podría explicar la conciencia, de modo que sería prematuro poner nuestras esperanzas sobre una versión demasiado particular que podría resultar imperfecta. (En el Apéndice B, me pondré en una situación difícil, sin embargo, en honor de todos los que quieren implicaciones verificables desde el principio.)

No son sólo las teorías de los filósofos las que precisan una puesta al día con un ejercicio de modelización a este nivel; las teorías de los investigadores del cerebro están en el mismo barco. Por ejemplo, la elaborada teoría de Gerald Edelman (1989) sobre los circuitos «de reentrada» del cerebro hace numerosas afirmaciones sobre el modo en que dichos circuitos pueden llevar a cabo tareas de discriminación, construir estructuras de memoria, coordinar la secuencia de la resolución de problemas y, en general, ejecutar las actividades de la mente humana, pero a pesar de la abundancia de detalles

neuroanatómicos y de las entusiastas y a menudo plausibles aseveraciones de Edelman, no sabremos si los circuitos de reentrada son la manera *correcta* de concebir la neuroanatomía funcional— hasta que no los hayamos forjado sobre una arquitectura cognitiva completa al nivel de descripción de ACT* o Soar, y hayan echado a andar.[1]

En un nivel de modelización más preciso, nos queda un asunto por resolver: el demostrar de qué manera se implementan las producciones (o comoquiera que llamemos a los demonios para el reconocimiento de patrones) en el cerebro. Baars (1988) denomina a sus especialistas «ladrillos» para la construcción, pero opta por dejar los detalles de la fabricación de ladrillos para otro día o para otra disciplina, aunque, como muchos han observado, resulta tentador suponer que los propios especialistas, a diferentes niveles de agregación, deberían modelarse como construcciones de tejidos *conexionistas* de uno u otro tipo.

El conexionismo (o PDP de *Parallel Distributed Processing*; «Procesamiento Paralelo Distribuido») es un reciente desarrollo dentro de la IA que promete aproximar la modelización cognitiva a la modelización neuronal, ya que los elementos que son *sus* ladrillos son nodos en redes paralelas que están conectados de manera que recuerda *bastante* a las redes neuronales del cerebro. El comparar la IA conexionista con la «Buena y Vieja IA» (Good Old Fashioned AI [GOFAI], Haugeland, 1985) y con muchos otros proyectos de modelización en el seno de las neurociencias se ha convertido en una de las mayores industrias en el mundo académico (véanse, por ejemplo, Graubard, 1988; Bechtel and Abrahamson, 1991; Ramsey, Stich y Rumelhart, 1991). Ello no debe sorprendernos, ya que el conexionismo enciende las primeras estelas remotamente plausibles conducentes a la unificación en la enorme *terra incognita* que se extiende entre las ciencias de la mente y las ciencias del cerebro. Sin embargo, casi ninguna de las polémicas en torno al «tratamiento adecuado del conexionismo» (Smolensky, 1988) afecta a nuestros proyectos aquí. *Es evidente* que deberá haber un nivel (o niveles) teórico(s) con la misma precisión descriptiva que los modelos conexionistas, y que este nivel mediará entre los niveles teóricos claramente neuroanatómicos y los niveles teóricos claramente psicológicos o cognitivos. La pregunta es qué ideas

1. Edelman (1989) es uno de esos teóricos que ha intentado tenerlo todo en cuenta, desde los detalles de la neuroanatomía a la psicología cognitiva, pasando por los modelos computacionales y las más abstrusas polémicas filosóficas. El resultado constituye un fracaso instructivo. Muestra con gran detalle de qué manera diferentes tipos de preguntas deben ser respondidas antes de poder afirmar que disponemos de una teoría completa de la conciencia, pero también muestra que ningún teórico puede apreciar las muchas sutilezas que presentan los diferentes problemas, en los diferentes campos. Edelman ha interpretado mal, y después ha rechazado con dureza, el trabajo de muchos aliados potenciales, de modo que ha aislado a su teoría del tipo de atención, informada y comprensiva, que necesita si quiere verla libre de sus errores y sus defectos. Estas observaciones traen a colación la cuestión paralela de que yo también podría haber subestimado algunos de los trabajos con los que me he mostrado en desacuerdo en estas páginas; sin duda así habrá sido, y espero que aquellos que hayan visto tratados injustamente a los retoños de su cerebro intentarán (otra vez) explicar aquello que yo no he entendido.

conexionistas serán parte de la solución y qué ideas se quedarán por el camino. En tanto que esta cuestión permanece sin resolver, los pensadores tienden a utilizar el ágora conexionista como un amplificador para lanzar sus eslóganes favoritos, y, aunque tengo tantos deseos como el que más de participar en este debate (Dennett, 1987b, 1988b, 1989, 1990c, 1991b,c,d), me voy a morder la lengua en esta ocasión y voy a continuar con mi tarea, que consiste en ver cómo podría surgir una teoría de la *conciencia* a partir de todo esto, cuando las aguas se hayan calmado, comoquiera que eso se produzca.

Nótese lo que ha ocurrido en la progresión desde la arquitectura de von Neumann hasta arquitecturas virtuales como los sistemas de producciones y (a un nivel más preciso) los sistemas conexionistas. Se ha producido lo que podríamos caracterizar como una alteración en el equilibrio del poder. Programas fijos y prediseñados, circulando por carriles con algunos ramales en función de los datos se han visto sustituidos por sistemas flexibles —y también volátiles—, cuya conducta ulterior está más en función de interacciones complejas entre lo que el sistema encuentra en este momento y lo que se ha encontrado en el pasado. En palabras de Newell, Rosenbloom y Laird (1989), «así, el problema con un ordenador tradicional es cómo interrumpirlo, mientras que el problema con Soar y ACT* (y, presumiblemente, también con la cognición humana) es cómo permanecer centrado» (pág. 119).

Con toda la tinta que se ha derramado por esta cuestión teórica, es importante hacer hincapié en el hecho de que se trata de una alteración en el *equilibrio* del poder, y no un cambio hacia un modo de operación «cualitativamente distinto». En lo más profundo del más volátil de los sistemas de reconocimiento de patrones («conexionistas» o no) subyace un motor de von Neumann, resoplando, computando una función computable. Desde el nacimiento de los ordenadores, los críticos de la inteligencia artificial no han cejado en sus ataques a la rigidez, la mecanicidad, el carácter *programado* de los ordenadores, y sus defensores han insistido repetidamente en afirmar que no se trata más que de un problema de grado de complejidad, que en un ordenador se pueden crear sistemas infinitamente dúctiles, difusos, holísticos, orgánicos. A medida que la IA se ha ido desarrollando, han ido apareciendo sistemas como éstos, de modo que los críticos ahora tienen que decidir si siguen pescando o si recogen el anzuelo: si declararan que los sistemas conexionistas, por ejemplo, son el tipo de cosa de la que siempre habían pensado que estaba hecha la mente, o si decidieran subir las apuestas e insistir en que para su gusto ni siquiera los sistemas conexionistas son lo bastante «holísticos», o lo bastante «intuitivos», o... (incluya aquí su eslogan favorito). Dos de los críticos de la IA más conocidos, los filósofos de Berkeley Hubert Dreyfus y John Searle, no se ponen de acuerdo sobre este punto: Dreyfus ha firmado una alianza con el conexionismo (Dreyfus y Dreyfus, 1988), mientras que Searle ha subido las apuestas al insistir que ningún ordenador conexionista podrá exhibir nunca una mentalidad *real* (1990a, 1990b).

Es posible que los escépticos «por principio» se estén batiendo en retirada, aunque los unificadores tienen todavía muchos problemas a los que en-

frentarse. El principal, en mi opinión, tiene una relación muy directa con nuestra teoría de la conciencia. El principal consenso en ciencia cognitiva, que podría ilustrarse con muchos diagramas como el de la figura 9.1, es que *por aquí* tenemos la memoria a largo plazo (el aviario de Platón) y que *por allá* tenemos el espacio o la memoria de trabajo, donde se produce el pensamiento.[2] Y sin embargo, no hay dos sitios en el cerebro que alberguen recursos separados. El único lugar en el cerebro capaz de albergar cualquiera de estas funciones separadas es todo el córtex; no dos lugares el uno al lado del otro, sino un único y gran espacio. Como señala Baars, resumiendo el amplio consenso, hay un espacio de trabajo *global*. Es global no sólo en el sentido funcional (dicho llanamente, es un «lugar» donde casi todo puede estar en contacto con casi todo lo demás), sino también en el sentido anatómico (está distribuido por todo el córtex, y no cabe duda de que también por otras regiones del cerebro). Ello significa, por tanto, que el espacio de trabajo debe valerse de las mismas regiones y redes neuronales que aparentemente juegan un papel fundamental en la memoria a largo plazo: el «almacenaje» de los cambios en el diseño fruto de la exploración individual.

Suponga que usted aprende a hacer pan de maíz, o que aprende el significado de «fenotípico». De un modo u otro, el córtex debe ser el medio en el que una serie de patrones de conexión estables fijan permanentemente esas alteraciones del diseño al cerebro con el que usted nació. Suponga que usted, de repente, se acuerda de su cita con el dentista, y que eso le quita todo el placer que le producía la música que estaba escuchando en ese momento. De un modo u otro, el córtex debe ser el medio en el que una serie

2. Los funcionalistas han convertido la «cajolología» en un hábito, dibujando diagramas que sitúan los diversos componentes funcionales en cajas separadas, al tiempo que niegan explícitamente que éstas tengan ninguna significación anatómica. (Yo mismo me confieso culpable de haber practicado, y de haber fomentado, esta práctica; véanse las figuras en *Brainstorms*, capítulos 7, 9 y 11.) Todavía pienso que «en principio» ésta es una buena táctica, aunque en la práctica tiende a privar al funcionalista de la visión de descomposiciones alternativas de las funciones y, en particular, de la idea de funciones múltiples superpuestas. La imagen de la separación espacial entre la memoria de trabajo y la memoria a largo plazo —una imagen tan antigua como el aviario de Platón— juega un papel nada baladí en la manera en que los teóricos interpretan las tareas cognitivas. Un ejemplo sorprendente: «La necesidad de los símbolos surge porque no es posible que toda la estructura que interviene en una computación esté ensamblada con anterioridad en la localización física donde se produce esa computación. Así pues, es necesario viajar hacia otras partes (distales) de la memoria a fin de obtener la estructura adicional» (Newell, Rosenbloom y Laird, 1989, pág. 105). Ello lleva directamente a la imagen de los *símbolos en movimiento*, y de ahí (especialmente en aquellos que son fervientes admiradores de esta idea), al escepticismo frente a las arquitecturas conexionistas, sobre la base de que los elementos que en estas arquitecturas se parecen más a los símbolos —los nodos que de una manera u otra fijan la semántica del sistema— son inamovibles dentro de su red de interconexiones. Véase, por ejemplo, Fodor y Pylyshyn (1988). El problema de los elementos semánticos fijos frente a los elementos semánticos en movimiento es una manera de observar un problema fundamental todavía no resuelto dentro de la ciencia cognitiva. Probablemente no es la mejor manera de abordarlo, pero no desaparecerá hasta que lo hayamos sustituido por una versión mejor, anclada en una aceptación positiva —por oposición a un rechazo histérico— de los hechos fundacionales de la neuroanatomía funcional.

de patrones de conexión inestables pueden alterar rápidamente esos contenidos transitorios de todo el «espacio», sin borrar, en el proceso, la memoria a largo plazo. ¿Cómo es posible que esos dos tipos tan diferentes de «representaciones» coexistan dentro del mismo medio y en el mismo tiempo? En los modelos exclusivamente cognitivos, las diversas tareas pueden representarse mediante cuadros separados en un diagrama, pero cuando tenemos que superponerlos a un único lienzo de tejidos neuronales, el simple problema del empaquetamiento es la menor de nuestras preocupaciones.

Se puede suponer que dos sistemas en red funcionalmente distintos estén dispuestos de tal modo que se interpenetren el uno con el otro (como el sistema de cableado telefónico y el sistema de autopistas que cubren todo el continente), pero éste no es el problema. El verdadero problema se halla justamente bajo la superficie de un presupuesto que hemos venido aceptando. Hemos supuesto, que cada demonio especialista se dedica a reclutar a otros cuando se presenta una tarea a gran escala. Si se tratara simplemente de llamar a filas a esos nuevos efectivos a fin de que pongan en práctica sus talentos *especializados* para una causa común, ya dispondríamos de modelos para estos procesos —como ACT*, Soar y el Espacio de Trabajo Global de Baars— con descripciones más o menos detalladas y más o menos plausibles. Pero, ¿y si a los especialistas se les reclutara en tanto que *generalistas*, a fin de contribuir a funciones en las que sus talentos especializados no juegan ningún papel en particular? Ésta es una idea que, por diversas razones (véase, por ejemplo, Kinsbourne y Hicks, 1978), resulta bastante atractiva, pero, por lo que yo sé, no existen todavía modelos computacionales sobre cómo podrían operar estos elementos con una doble función.

Éste es el problema: se suele suponer que los especialistas del cerebro deben adquirir su identidad funcional en virtud de su posición en una red de conexiones más o menos fijas. Por ejemplo, parece que los únicos hechos capaces de explicar «el interés por» el color de una determinada área neuronal sean aquellos que tienen que ver con sus conexiones idiosincrásicas, por indirectas que sean, con las células cónicas de la retina que son las más sensibles a las diferencias de frecuencia de la luz. Una vez que se ha establecido una identidad funcional de este tipo, se podrían cortar estas conexiones (como ocurre con las personas que quedan ciegas durante la edad adulta), sin que por ello desapareciera (por completo) la capacidad de los especialistas para representar (o, de un modo u otro «interesarse por») el color, aunque sin estas conexiones causales, resulta difícil ver qué podría asignar a los especialistas un papel con contenido específico.[3] Parece, por tanto, que el córtex se compone (en gran medida) de elementos cuyos poderes de representación más o menos fijos son el resultado de su localización funcional dentro del conjunto de la red. Representan algo de la misma manera que los miembros del Parlamento representan a una provincia: transportando

3. En otras palabras, los atractivos de una «teoría causal de la referencia» son tan evidentes para los investigadores en ciencia cognitiva como para los filósofos.

portando información desde las fuentes a las que están específicamente ligados (por ejemplo, la mayoría de sus conversaciones telefónicas desde su escaño se producen con un punto u otro de su provincia). Imagine ahora a los miembros del Parlamento sentados todos juntos en un estadio y representando el importante mensaje «¡La velocidad mata!», sosteniendo unos cartones de color sobre sus cabezas que forman las letras gigantes del mensaje, visible desde el otro lado del estadio. En pocas palabras, *píxels* vivientes, cuya relación con sus circunscripciones no juega ningún papel en su contribución a la representación del grupo. Algunos modelos de reclutamiento cortical incorporan la idea de que *algo parecido* a este papel representacional secundario debe de ser posible. Por ejemplo, resulta tentador suponer que el contenido informativo sobre un determinado asunto pueda surgir en un área especializada determinada y que después, de un modo u otro, sea propagado a través de las regiones corticales, explotando la variabilidad de estas regiones sin comprometer la semántica especializada de las unidades que residen en esa área. Supóngase, por ejemplo, que se produce un cambio repentino en el cuadrante superior izquierdo del mundo visual de una persona. Como es de esperar, el surgimiento de la excitación cerebral se percibe en primer lugar en aquellas partes del córtex visual que representan (a la manera parlamentaria) las diversas propiedades de los acontecimientos en el cuadrante superior izquierdo de la visión, pero esos puntos calientes se convierten inmediatamente en las fuentes de la puesta en marcha de un proceso de propagación que va involucrando a nuevos agentes corticales con circunscripciones distintas. Si esta propagación del despertar por las áreas del córtex no es una mera fuga o un simple ruido, si juega algún papel crucial al elaborar o al facilitar la corrección del borrador de un fragmento narrativo, los agentes reclutados deben jugar un papel totalmente distinto del que desempeñan cuando son la fuente original.[4]

No debe sorprendernos que no dispongamos, todavía, de buenos modelos de dicha funcionalidad múltiple (las únicas especulaciones plausibles que conozco son las de Minsky en *The Society of Mind*). Como señalamos en el capítulo 7, los ingenieros humanos, con su imperfecta capacidad de previsión, se afanan en diseñar sistemas en los que cada elemento juega un único papel, cuidadosamente aislado para evitar interferencias del exterior, a fin de minimizar las devastadoras consecuencias de los efectos secundarios no previstos. La Madre Naturaleza, por otra parte, no se preocupa por los efectos secundarios no previstos y así, puede capitalizar aquellos efectos secundarios felices que ocasionalmente (de uvas a peras) se producen. Es probable que la inescrutabilidad de la descomposición funcional del córtex, que hasta ahora ha mantenido en jaque a los investigadores del cerebro, resulte del hecho de que éstos se sientan constitucionalmente incapacitados para defender hipótesis en las que se atribuyen papeles múltiples a los ele-

4. Fodor nota una variante de este problema en su discusión de «alojar un concepto» (1990, págs. 80-81).

mentos disponibles. Algunos románticos —el filósofo Owen Flanagan (1991) los denomina nuevos misteriales *(New Mysterians)*— han avanzado la idea de que el cerebro se encuentra con una barrera insuperable en el momento de comprender su propia organización (Nagel, 1986, y McGinn, 1990). Yo no estoy defendiendo esta idea, me limito a señalar que llegar a comprender cómo funciona el cerebro se está mostrando como una empresa endiabladamente difícil —aunque no imposible—, en parte porque éste fue diseñado por un proceso capaz de prosperar en un medio de funcionalidad múltiple y superpuesta, algo muy complicado de discernir desde la perspectiva de la ingeniería inversa.

Estos problemas, en el caso de que hayan sido percibidos, no provocan más que vagas e ilusionadas alusiones. Algunos sienten la tentación de rechazar sin más la idea de esta dualidad especialista/generalista, y no porque puedan probar que es errónea, sino porque no pueden imaginar cómo modelarla y, por tanto, albergan la razonable esperanza de que nunca tendrán que hacerlo. Pero una vez se ha abierto la perspectiva, ésta ofrece, cuando menos, nuevas pistas sobre dónde buscar. Los neurofisiólogos han identificado (provisionalmente) mecanismos en las neuronas como los receptores NMDA y las sinapsis de von der Malsburg (1985), que son posibles candidatos a cumplir la función de moduladores de la conectividad entre células. Dichas puertas podrían permitir la rápida formación de «ensamblajes» transitorios, que podrían a su vez superponerse a las redes sin el requisito de que se produzcan alteraciones de las fuerzas sinápticas a largo plazo que suelen ser consideradas como la cola que mantiene unidos los ensamblajes permanentes de la memoria a largo plazo. (Para algunas especulaciones recientes en esta misma línea, véase Flohr, 1990.)

A mayor escala, los neuroanatomistas han seguido completando el mapa de las conexiones del cerebro, mostrando no sólo qué áreas permanecen activas en qué circunstancias, sino mostrando también aunque en menor medida, cuáles son las contribuciones de cada área. Existen numerosas hipótesis sobre el papel que jugarían diversas áreas en la conciencia. La *formación reticular* en el diencéfalo y el *tálamo* por encima de ella, hace tiempo que son conocidos por el papel que juegan en la tarea de despertar el cerebro —del sueño, por ejemplo, o en respuesta a una novedad o una emergencia—, y ahora que las vías de conexión están más claramente dibujadas, es posible formular y verificar nuevas y más detalladas hipótesis. Crick (1984), por ejemplo, propone que los ramales que nacen en el tálamo y se dirigen hacia todas las partes del córtex encajan perfectamente en el papel de un «reflector», despertando o ensalzando diferencialmente ciertas áreas especializadas y reclutándolas para los objetivos del momento.[5] Baars (1988) ha elaborado

5. Las teorías de la atención basadas en la idea del reflector han sido muy populares durante años. Las teorías más toscas cometen el error de ser demasiado literales al suponer que lo que el reflector ilumina o ensalza en un determinado momento es una región de *espacio visual*: exactamente igual que un foco de teatro que ilumina una región del escenario. Las teorías del reflector más defendibles, aunque por el momento también las más impresionistas, insisten en seña-

una idea similar: el ERTAS o Extended Reticular Thalamic Activating System (Sistema de Activación Talámico-Reticular Extendido). Sería bastante fácil incorporar dicha hipótesis en nuestro enfoque no comprometido con la anatomía de la competición entre coaliciones de especialistas, siempre que no caigamos en la tentación de imaginar un jefe talámico que comprende los acontecimientos regidos por las diversas partes del cerebro con las cuales está «en comunicación».

Similarmente, los *lóbulos frontales* del córtex, la parte del cerebro que más ha crecido en el *Homo sapiens*, son conocidos por su participación en el control a largo plazo, y en la organización y en la secuenciación de la conducta. Las lesiones en diversas regiones de los lóbulos frontales producen generalmente síntomas contrapuestos tales como la facilidad para distraerse frente a la total incapacidad, por exceso de atención, de abandonar una tarea, y la impulsividad frente a la incapacidad de seguir esquemas de acción que requieren una gratificación retardada. Resulta, pues, tentador instalar al jefe en los lóbulos frontales, y muchos modelos se han movido en esta dirección. Un modelo particularmente elaborado es el Sistema de Supervisión de la Atención, de Norman y Shallice (1985), que ellos localizan en el córtex prefrontal y al que conceden la particular responsabilidad de la resolución de conflictos cuando las burocracias subsidiarias no pueden cooperar. De nuevo, hallar una localización anatómica para los procesos que son básicos en el control de lo que ocurrirá después es una cosa, mientras que localizar al jefe es otra cosa muy distinta; cualquiera que vaya a la caza de la pantalla frontal donde el jefe lleva la cuenta de los procesos que está controlando se encontrará persiguiendo sombras (Fuster, 1981; Calvin, 1989a).

Sin embargo, una vez hemos abjurado de todas estas imágenes tan tentadoras, tenemos que encontrar nuevas maneras de pensar en las contribuciones que están haciendo estas áreas, y aquí estamos todavía muy faltos de ideas, a pesar de los muchos progresos que hemos realizado. El problema no se debe tanto al hecho de que desconozcamos por completo la maquinaria, sino al hecho de que carecemos de un modelo computacional de lo que hace esa maquinaria, y de cómo lo hace. Aquí todavía estamos en la etapa de las metáforas y de las vagas alusiones, pero no es una etapa que debamos evitar; es una etapa por la que hay que pasar en nuestro camino hacia el desarrollo de modelos más explícitos.

4. Los poderes de la máquina joyceana

De acuerdo con nuestro esbozo, en el cerebro se produce una competición entre diversos acontecimientos llenos de contenido, de los cuales sola-

---

lar que es una porción de *espacio conceptual* o *semántico* la que se ve ensalzada (imagine usted, si puede, un foco de teatro capaz de iluminar solamente a los *Capuletos*, o solamente a los *amantes*). Véase Allport (1989) para las dificultades que presentan las teorías del reflector.

mente un selecto conjunto es el que se proclama «vencedor». Es decir, consiguen engendrar diferentes tipos de efectos continuados. Algunos, al aunar diversos demonios del lenguaje, contribuyen a ulteriores actos de enunciación, tanto enunciados en voz alta dirigidos a otros como enunciados silenciosos dirigidos a uno mismo. Otros prestan su contenido a otras formas subsiguientes de autoestimulación tales como dibujar para uno mismo. Los demás mueren casi de inmediato, dejando sólo leves huellas —evidencias circunstanciales— de su existencia pasada. Es posible que usted se pregunte qué tiene de bueno que algunos contenidos consigan, de este modo, el acceso a ese círculo encantado, y qué es lo que convierte en encantado a *este* círculo. Supuestamente, la conciencia es algo muy especial. ¿Qué tiene de especial el ganar el acceso a la siguiente ronda del ciclo de autoestimulación? ¿De qué manera presta ayuda? ¿Acaso los poderes mágicos se corresponden con los acontecimientos que se producen en estos mecanismos?

He procurado evitar cualquier afirmación que indique que una victoria en este remolino competitivo equivalga a una elevación hacia la conciencia. Sí que he procurado insistir, en cambio, en el hecho de que no existe ningún motivo para trazar una línea que separe los acontecimientos que están claramente «en» la conciencia de los acontencimientos que siempre quedarán «fuera» o «por debajo» de la conciencia. (Véase Allport, 1988, donde se presentan nuevos argumentos en favor de esta posición.) No obstante, si mi teoría de la máquina joyceana está llamada a aportar nueva luz sobre el problema de la conciencia, sería bueno que algunas de las actividades de esta máquina, si no todas, poseyera algún rasgo extraordinario, ya que es innegable que la conciencia es, intuitivamente, algo muy especial.

Es difícil enfrentarse a estas cuestiones sin caer en la trampa de pensar que en primer lugar debemos entender *para qué* sirve la conciencia, a fin de poder preguntarnos después si los mecanismos propuestos tendrían éxito en llevar a cabo *esa* función, cualquiera que ésta sea.

En su influyente libro *Vision* (1982), el investigador del cerebro y de la IA David Marr propuso tres niveles de análisis que deberían ser tomados en consideración por todo intento de explicar cualquier fenómeno mental. El nivel más abstracto o «superior», el nivel computacional, comporta un análisis de «*el problema* (la cursiva es mía) en tanto que tarea de procesamiento de la información», mientras que el nivel intermedio, el algorítmico, comporta un análisis de los procesos reales a través de los cuales *dicha* tarea de procesamiento de la información se lleva a cabo. El nivel más bajo, el nivel físico, comporta un análisis de la maquinaria neuronal y muestra de qué manera ésta ejecuta los algoritmos descritos en el nivel intermedio, realizando así su tarea tal como se describe de forma abstracta en el nivel computacional.[6]

6. Distinciones próximas a ésta son mi trinidad formada por la estrategia intencional, la estrategia del diseño y la estrategia física (Dennett, 1971), y la identificación por parte de Allen Newell (1982) del «nivel del conocimiento» por encima del «nivel del sistema de símbolos físicos». Véanse Dennett (1987a, 1988e), y Newell (1988).

Los tres niveles de Marr también podrían utilizarse para describir cosas que son mucho más simples que las mentes. Precisamente, podemos hacernos una idea de cuáles son las diferencias ente ellos viendo cómo se aplican a algo tan simple como un ábaco. Su tarea computacional es llevar a cabo cálculos aritméticos: producir una solución correcta a un problema aritmético, dados unos datos iniciales. A este nivel, pues, un ábaco y una calculadora de bolsillo son iguales; han sido diseñados para efectuar la misma tarea de «procesamiento de la información». La descripción algorítmica del ábaco es lo que usted aprende cuando aprende cómo manipularlo, es decir, las reglas para mover las cuentas durante el acto de sumar, restar, multiplicar o dividir. Su descripción física depende de lo que esté hecho: podría ser de cuentas de madera enhebradas en varillas sostenidas por un marco, o podría estar hecho con fichas de póker alineadas a lo largo de las juntas de las baldosas del suelo, o podría hacerse con un lápiz y una goma sobre una hoja de papel cuadriculado.

Marr hizo la recomendación de modelar los fenómenos psicológicos a los tres niveles de análisis, aplicando un particular hincapié en la importancia de ser muy claro en todo lo referente al nivel superior, el computacional, antes de apresurarse a modelar los niveles inferiores.[7] En su propio trabajo sobre la visión, Marr demostró el poder de esta estrategia, y, desde entonces, otros investigadores la han puesto en práctica con otros fenómenos. Resulta tentador aplicar los tres mismos niveles de análisis a la conciencia en bloque, y algunos han caído en esa tentación. Sin embargo, como vimos en el capítulo 7, ésta es una simplificación muy arriesgada: al preguntarnos «¿cuál es la función de la conciencia?» estamos asumiendo que hay una única «tarea de procesamiento de la información» (por muy compleja que sea) para la cual la maquinaria neuronal de la conciencia está diseñada —presumiblemente, por la evolución—. Ello nos puede llevar a pasar por alto otras posibilidades importantes: que algunos rasgos de la conciencia posean funciones múltiples; que algunas funciones de la conciencia estén mal servidas por los rasgos existentes, debido a limitaciones históricas sobre su desarrollo; que algunos rasgos no tengan ninguna función o, cuando menos, ninguna función que *nos* reporte algún beneficio. Cuidando de evitar estas equivocaciones, pues, pasemos a examinar los *poderes* (y no necesariamente las funciones) de los mecanismos descritos en mi esbozo diminuto.

En primer lugar, y como pudimos comprobar en el capítulo 7, existen problemas significativos de autocontrol creados por la proliferación de especialistas activos simultáneamente, y una de las tareas fundamentales llevada a cabo por las actividades de la máquina joyceana es resolver dispu-

7. Como señaló Marr, «Al separar la explicación en diferentes niveles, se abre la posibilidad de hacer afirmaciones explícitas sobre qué se está computando y por qué, y construir teorías capaces de declarar que lo que se computa es, en cierto sentido, óptimo o que ofrece la garantía de un funcionamiento correcto» (pág. 19). En Dennett (1971, 1983, 1987a, 1988d) se presentan nuevos argumentos sobre las ventajas y los inconvenientes de esta ingeniería inversa; véase también Ramachandran (1985).

tas, suavizar la transición de un régimen a otro e impedir inoportunos *golpes de Estado* organizando las fuerzas «apropiadas». Las tareas simples o convertidas en rutinarias gracias a un aprendizaje repetido y sin competición pueden ejecutarse sin necesidad de alistar nuevos efectivos y, por tanto, son inconscientes, pero cuando una tarea es difícil o desagradable requiere «concentración», que «nosotros» podemos obtener con la ayuda de actos de autoamonestación y muchos otros trucos mnemotécnicos, ensayos (Margolis, 1989), y otros actos de automanipulación (Norman y Shallice, 1985). A veces descubrimos que ayuda hablar en voz alta, una vuelta atrás hacia esas burdas pero efectivas estrategias cuyos dignos descendientes son nuestros pensamientos privados.

Dichas estrategias de autocontrol nos permiten gobernar nuestros propios procesos perceptivos de maneras que abren nuevas oportunidades. Como ha señalado el psicólogo Jeremy Wolfe (1990), nuestros sistemas visuales poseen un diseño innato que les permite detectar ciertos tipos de cosas —aquellas cosas que «aparecen» cuando «miramos»—, pero existen otros tipos de cosas que sólo podemos identificar si podemos *buscarlos*, deliberadamente, en el marco de un protocolo establecido por un acto de autorrepresentación. Un punto rojo entre un montón de puntos verdes resaltará tanto como una mosca en la leche (de hecho, resaltará como una cereza madura entre las hojas de un árbol), pero si su proyecto consiste en hallar un punto rojo entre una multitud de puntos multicolores, usted tendrá que fijarse *a sí mismo* una tarea de búsqueda serial. Y si su proyecto consiste en hallar un confeti rojo cuadrado en un montón de confetis de muchas otras formas y colores (o responder a la pregunta «¿dónde está Wally?» [Hanford, 1987] de los populares dibujos-enigma), la tarea de búsqueda serial puede convertirse en un proyecto particularmente absorbente, necesitado de un cierto método y un alto grado de autocontrol.

Estas técnicas para representarnos cosas a nosotros mismos nos permiten erigirnos en rectores o ejecutivos de una manera a la que ninguna otra criatura puede aproximarse. Podemos elaborar protocolos con antelación, gracias a nuestra capacidad para el pensamiento hipotético y el cambio de escenarios; podemos reforzar nuestro propósito de participar en proyectos desagradables o a largo plazo mediante hábitos de rememoración, y ensayando los beneficios y los costes esperados de las políticas que hayamos adoptado. Aún más importante, esta práctica del ensayo crea un recuerdo de la ruta por la que hemos llegado adonde estamos (lo que los psicólogos denominan la memoria *episódica*), de modo que podemos explicarnos, cuando nos encontramos acorralados, qué errores hemos cometido (Perlis, 1991). En el capítulo 7 vimos de qué modo el desarrollo de estas estrategias permitió a nuestros antepasados mirar hacia el futuro, y parte de lo que les dotó de esta mejorada capacidad de anticipación fue una mejorada capacidad para el recuerdo: el ser capaces de mirar hacia atrás, hacia sus acciones recientes, para ver dónde cometieron errores. «¡Vaya, no debería volver a hacer *esto!*» son las palabras de cualquier criatura que aprende de la experiencia,

pero nosotros podemos aprender a proyectar nuestros *estos* mucho más hacia atrás y con mucha más lucidez que cualquier otra criatura, gracias a nuestro hábito de tomar nota de lo ocurrido o, más exactamente, gracias a nuestros hábitos de autoestimulación que tienen, entre muchos otros efectos, el de perfeccionar el recuerdo.

Sin embargo, este recargar de la memoria es sólo uno de los muchos efectos beneficiosos de estos hábitos. Igual de importante es el efecto de transmisión (Baars, 1988), que crea una especie de foro, que permite a *cualquier* cosa que uno ha aprendido contribuir a *cualquier* problema en curso. Baars desarrolla la idea de que esta mutua accesibilidad de los contenidos proporciona el contexto sin el cual los acontecimientos que se producen «en la conciencia» no tendrían —*o no podrían tener*— sentido para el sujeto. Los contenidos que componen el contexto circundante no son siempre conscientes —de hecho, no son accesibles, a pesar de haber sido activados—, pero las conexiones entre ellos y los contenidos que pueden surgir en los testimonios verbales es lo que fija aquello que podríamos denominar su significado «conscientemente aprehendido».

En una línea similar, Ray Jackendoff (1987) argumenta que los niveles superiores de análisis llevados a cabo por el cerebro, palabras con las que Jackendoff se refiere a los niveles más abstractos, *no* son accesibles a la experiencia, pese a ser los que hacen posible esa experiencia, confiriéndole significado. Su análisis nos proporciona, pues, un útil antídoto contra otra de las encarnaciones del Teatro Cartesiano en tanto que «cumbre» o «punta del iceberg». (He aquí un buen ejemplo, debido al neuropsicólogo Roger Sperry: «En una posición de alto mando en los niveles superiores de la jerarquía de organización cerebral, las propiedades subjetivas (...) ejercen un control sobre las actividades biofísicas y químicas de los niveles subordinados» [1977, pág. 117].)

Bastantes filósofos, principalmente los influidos por la escuela husserliana de la fenomenología (Dreyfus, 1979; Searle, 1983), han hecho hincapié en la importancia de ese «trasfondo» de la experiencia consciente, pero han mostrado una tendencia a describirlo como un rasgo misterioso e intratable, que desafía toda explicación mecanicista, en vez de describirlo, según han propuesto Baars y Jackendoff, como la clave que permitirá desarrollar una teoría computacional de lo que ocurre. Estos filósofos han asumido que la conciencia es el origen de un tipo especial de «intencionalidad intrínseca», aunque, como ha señalado el filósofo Robert van Gulick, ello no tiene otro efecto que el de hacernos retroceder.

> La experiencia de la comprensión a nivel personal... no es una ilusión. Yo, el sujeto personal de la experiencia, comprendo. Yo puedo llevar a cabo todas las conexiones necesarias dentro de la experiencia, apelando a las representaciones para conectarlas inmediatamente las unas con las otras. El hecho de que mi capacidad sea el resultado de que yo esté compuesto por un sistema organizado de componentes subpersonales que producen mi flujo ordenado

> de pensamientos no impugna mi capacidad. Lo que es ilusorio o erróneo es únicamente la visión de que yo soy un yo sustancial distinto que produce estas conexiones en virtud de una forma de comprensión totalmente ajena a la conducta (van Gulick, 1988, pág. 96).

Cualquiera de las cosas que usted ha aprendido puede contribuir a cualquiera de las cosas a las que se está enfrentando. Al menos ésta es la imagen ideal. Esta propiedad fue bautizada con el nombre de *isotropía* por Fodor (1983), el poder, como diría Platón, de hacer que acudan, o por lo menos que canten, los pájaros pertinentes cuando se les necesita. Parece mágico, pero como sabe todo mago de teatro, la apariencia mágica se ve potenciada por el hecho de que se suele poder contar con la audiencia para exagerar un fenómeno que necesita de una explicación. En un principio podemos parecer idealmente isotrópicos, pero no lo somos. Una reflexión más pausada nos traerá a la mente todas aquellas ocasiones en que no hemos atinado a reconocer la significación de los datos a tiempo. Piénsese en la exageración de las comedias clásicas: la «toma doble» (Neisser, 1988). A veces, incluso nosotros serramos la rama sobre la que estamos sentados o encendemos una cerilla para comprobar el nivel del depósito de gasolina.[8]

Los magos saben que, a veces, un conjunto de trucos fáciles basta para hacer «magia», y también lo sabe la Madre Naturaleza, el último creador de artilugios. Los investigadores en inteligencia artificial han estado explorando el espacio de los trucos posibles, en busca de «un manojo de... heurísticas adecuadamente coordinadas y rápidamente desplegadas» (Fodor, 1983, pág. 116) que pudieran proporcionar el punto justo de isotropía que poseen los pensadores humanos. Modelos como ACT* y Soar —y muchas otras visiones desarrolladas en el seno de la IA— resultan prometedores pero poco concluyentes. Algunos filósofos, principalmente Dreyfus, Searle, Fodor y Putnam (1988), están convencidos de que esta idea de la mente como un aparato es equivocada, y han intentado construir argumentos que prueben la imposibilidad de esta empresa (Dennett, 1988b, 1991c). Fodor, por ejemplo, observa que mientras los sistemas con un objeto específico pueden estar preconfigurados en un sistema con funciones más generales, capaz de responder con versatilidad a cualquier elemento que se presente, «lo que cuenta puede ser la conectividad *inestable e instantánea*» (pág. 118). Duda de que alguien llegue a desarrollar algún día una teoría sobre esta conectividad, pero la suya no es una mera visión pesimista, Fodor duda *por principio* (un buen truco). Está en lo cierto al afirmar que debemos esperar que nuestra aproximación a la isotropía se deba a nuestro *software*, y no a nuestra configuración inicial, pero su argumento en contra de la hipótesis del «maletín de trucos» presupone que somos mejores «tomándolo todo en consideración» de lo que

8. En *Minimal Rationality* (1986), el filósofo Christopher Cherniak analiza las perspectivas y las limitaciones de procesos deductivos como los que posibilitan la existencia de un foro abierto. Véase también Stalnaker (1984).

realmente somos. Somos buenos, pero no fantásticos. Los hábitos de autoestimulación que desarrollamos nos convierten en astutos explotadores de esos recursos que tanto nos ha costado conseguir; no siempre conseguimos que cante el pájaro que queríamos que cantara en ese momento, pero lo conseguimos las veces suficientes como para sentir su agradable compañía.

5. Pero, ¿es esto una teoría de la conciencia?

Hasta ahora he venido manteniendo una actitud bastante reservada en relación a la conciencia. Con mucho cuidado he evitado contarles lo que dice mi teoría sobre qué *es* la conciencia. No he afirmado que cualquier cosa que «instancie» una máquina joyceana sea consciente, ni he afirmado que ningún estado en particular de dicha máquina virtual sea un estado consciente. Las motivaciones de mi reticencia eran tácticas: quería evitar disputas sobre qué es la conciencia hasta que hubiera tenido una oportunidad de demostrar que al menos un buen número de los presuntos poderes de la conciencia podían explicarse mediante los poderes de la máquina joyceana *tanto* si una máquina joyceana dota de conciencia al *hardware* que la alberga *como si no.*

¿Podría acaso existir un ser no consciente con un espacio de trabajo global interno en el que unos demonios transmiten mensajes para otros demonios, formando coaliciones y todo lo demás? Si así fuera, entonces la sorprendente capacidad humana para ajustar con versatilidad y rapidez los estados mentales como respuesta a casi cualquier contingencia, por novedosa que sea, nada debería a la conciencia, sino simplemente la arquitectura computacional que hace posible esta intercomunicación. Si la conciencia es algo además de la máquina joyceana, entonces todavía no he presentado una teoría de la conciencia, aunque haya despejado algunas misteriosas incógnitas.

Hasta que el esbozo de la teoría no estuvo completamente ensamblado, he tenido que diferir la aclaración de todas estas dudas pero, finalmente, ha llegado el momento de coger el toro por los cuernos y de enfrentarme a la conciencia, ese maravilloso misterio. Y ahora declaro que SÍ, que mi teoría es una teoría de la conciencia. Cualquiera o cualquier cosa que posea dicha máquina virtual como su sistema de control es consciente en todos los sentidos, y es consciente *porque* posee esa máquina virtual.[9]

Ahora estoy preparado para dar cuenta de todas las objeciones que la

9. Jackendoff (1987) adopta una táctica ligeramente distinta. Divide el problema de la mente y el cuerpo en dos subproblemas, y dirige su teoría hacia la pregunta de cómo encaja en el cuerpo la mente *computacional*; de este modo, le queda por resolver un «problema de la mente y la mente», es decir, cuál es la relación entre la mente fenomenológica y la mente computacional. En lugar de aceptar esto como un misterio residual, yo me propongo demostrar de qué modo el modelo de las Versiones Múltiples, junto al método de la heterofenomenología, elimina ambos problemas a la vez.

teoría pudiera suscitar. ¿Podría una entidad no consciente —un zombie, por ejemplo— poseer una máquina joyceana? Esta pregunta plantea una objeción que es tan común en momentos como éste, que el filósofo Peter Bieri (1990) la ha bautizado con el nombre de *La Rueda de Oraciones Tibetana.* Siempre vuelve, una y otra vez, independientemente de cuál sea la teoría propuesta:

> Todo esto está muy bien; todos esos detalles funcionales sobre cómo el cerebro hace esto y lo otro, pero puedo imaginarme *todo esto* ocurriendo en una entidad, ¡sin que haya ninguna conciencia real!

Una buena respuesta para estas observaciones, que rara vez se da, es: ¿Ah sí? ¿Y cómo lo sabe usted? ¿Cómo sabe usted que ha imaginado «todo esto» con el suficiente detalle y con la suficiente atención hacia todas sus implicaciones? ¿Qué le hace pensar que su afirmación es una premisa conducente a alguna conclusión interesante? Pensemos en lo fríos que nos dejaría un *vitalista* moderno si dijera:

> Todo esto está muy bien; todos esos detalles sobre el ADN, las proteínas y todo lo demás, pero puedo imaginar la posibilidad de descubrir una entidad que fuera y actuara como un gato, de la sangre que circulase por sus venas al ADN de sus «células», pero que no estuviera vivo. (¿Puedo realmente? Por supuesto: ahí está, maullando, y después Dios me murmura al oído, «¡no está vivo! ¡No es más que un no-sé-qué mecánico hecho de ADN!». Y yo, en mi imaginación, Le creo.)

Confío en que nadie piense que éste es un buen argumento en favor del vitalismo. Ese esfuerzo de imaginación no cuenta. ¿Por qué no? Porque es un retruécano demasiado débil para ser comparado con la explicación de la vida que nos presenta la biología contemporánea. Lo único que nos demuestra este «argumento» es que podemos ignorar «todo eso» y aferrarnos a la convicción de que estamos decididos a hacerlo. ¿Es la Rueda de Oraciones Tibetana mejor argumento que la teoría que he esbozado?

Gracias a los ejercicios de imaginación llevados a cabo en los capítulos precedentes, nos hallamos en posición de traspasar el peso de la prueba. La Rueda de Oraciones Tibetana (que, como veremos, tiene muchas e importantes variantes distintas) es un descendiente del famoso argumento de Descartes (véase el capítulo 2) en el que afirma hallarse en disposición de concebir *clara y distintamente* que su mente es diferente de su cerebro. La fuerza de dicho argumento depende por completo de lo alto que ponga uno el listón del acto de concebir. Puede que algunos afirmen que pueden concebir clara y distintamente cuál es el mayor número primo o cómo sería un triángulo que no fuese una figura rígida. Están equivocados o, en cualquier caso, lo que sea que estén haciendo cuando dicen concebir estas cosas no debería interpretarse como un signo de lo que es posible. Ahora estamos en situación de imaginarnos «todo eso» con cierto detalle. ¿Puede usted *realmente*

imaginarse un zombie? El único sentido en que es «obvio» que usted puede no es un sentido que constituya una amenaza para mi teoría, y la adopción de un sentido más fuerte, quizá menos obvio, requiere la demostración de que es efectivamente posible.

Por lo general, los filósofos no lo han exigido. Los experimentos mentales más influyentes en la filosofía de la mente contemporánea comportan, todos, una invitación a la audiencia a imaginar alguna situación especialmente tramada o estipulada para la ocasión, seguida —sin controlar si dicha proeza imaginativa efectivamente se llevó a cabo correctamente— por una invitación a «observar» las diferentes consecuencias de tal fantasía. Estas «bombas de intuición», como yo las llamo, son, a veces, recursos endiabladamente inteligentes. Se merecen la fama que tienen aunque sea solamente por su capacidad de seducción.

En la tercera parte, daremos buena cuenta de todos ellos, y desarrollaremos una teoría de la conciencia a medida que avanzamos. Desde nuestra nueva perspectiva, seremos capaces de percibir el juego de manos que engaña a la audiencia —y a los ilusionistas—, y, en el proceso, iremos perfilando nuestros propios poderes de imaginación. Entre los más famosos argumentos con los que nos encontraremos no está solamente la presunta posibilidad de que haya zombíes, sino también el espectro invertido, lo que Mary, la investigadora del color, no sabe sobre los colores, la habitación china, y lo que se siente al ser un murciélago.

TERCERA PARTE

LOS PROBLEMAS FILOSÓFICOS DE LA CONCIENCIA

Capítulo 10

MOSTRAR Y CONTAR

1. Rotando imágenes en el ojo de la mente

El primer desafío que deberemos afrontar, antes de enfrentarnos a los experimentos mentales filosóficos, nos los presentan ciertos experimentos reales que parecerían tener la capacidad de rehabilitar al Teatro Cartesiano. Algunas de las investigaciones más interesantes en el ámbito de la ciencia cognitiva desarrolladas en los últimos veinte años tratan sobre la capacidad de los humanos para manipular imágenes mentales, y fueron iniciadas por el clásico estudio del psicólogo Roger Shepard (Shepard y Metzler, 1971) sobre la *velocidad de rotación mental* de imágenes como éstas.

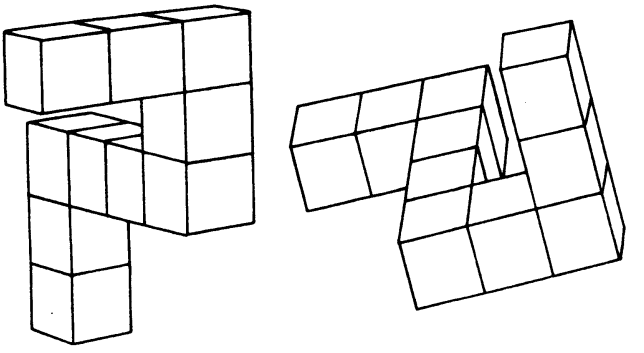


Figura 10.1

En el experimento original se mostraba a los sujetos pares de dibujos lineales como los de la figura y se les preguntaba si eran o no perspectivas diferentes de una misma forma. En este caso, como usted mismo puede comprobar, la respuesta es sí. ¿Cómo lo hizo? La respuesta que se suele dar es: «hice rotar una de las imágenes en mi ojo de la mente, y superpuse la una sobre la otra». Shepard hacía variar las distancias de rotación angular entre pares de figuras —algunos pares sólo estaban a algunos grados del eje de alineación, otros necesitaban rotaciones más largas para alinearse— y

medía el tiempo que se tomaban los sujetos para responder, en término medio, a diferentes presentaciones. En el supuesto de que en el cerebro se produjera algo parecido a un proceso real de rotación, la duración de este proceso debería de ser el doble en el caso de una rotación de 90 grados que en una rotación de 45 grados (ignorando la aceleración y la deceleración, y manteniendo constante la velocidad de rotación).[1] Los datos de Shepard confirman bastante bien esta hipótesis bajo una gran variedad de condiciones. Cientos de experimentos, realizados por Shepard y otros investigadores, se han dedicado al examen detallado del comportamiento de los mecanismos cerebrales para la manipulación de imágenes, y —por expresar el relativo consenso existente con toda la cautela que merece— parece que en el cerebro hay lo que el psicólogo Stephen Kosslyn (1980) denomina un «*buffer* visual», que ejecuta las transformaciones mediante procesos que son fuertemente «figurativos» o, por utilizar la expresión del propio Kosslyn, *cuasipictóricos*.

¿Qué significa esto? ¿Acaso han descubierto los investigadores en ciencia cognitiva que, a pesar de todo, el Teatro Cartesiano existe? Según Kosslyn, estos experimentos demuestran que las imágenes se ensamblan para la presentación interna de una manera que recuerda mucho a cómo pueden crearse imágenes en un TRC (un tubo de rayos catódicos como el de la pantalla de un televisor o de un ordenador) a partir de archivos en la memoria de un ordenador. Una vez están en la pantalla interna, los sujetos a los que se les solicita que efectúen algún tipo de tarea las pueden hacer rotar, las pueden buscar y las pueden someter a muchas otras manipulaciones. Kosslyn insiste, sin embargo, en el hecho de que su modelo del TRC es una metáfora. Lo cual debería recordarnos los talentos para la «manipulación de imágenes» de Shakey, y no cabe duda de que Shakey no tenía ningún Teatro Cartesiano en su ordenador-cerebro. A fin de tener una idea más clara de lo que realmente debe de estar ocurriendo en el cerebro humano, podemos empezar por un modelo *no* metafórico, demasiado poco restrictivo para ser válido y después ir «sustrayendo» las propiedades que no nos parezcan deseables una por una. En otras palabras, tomaremos el modelo del TRC de Kosslyn y, gradualmente, le iremos añadiendo restricciones.

Considérese, en primer lugar, un sistema que realmente manipula imágenes como los sistemas de gráficos por ordenador que hoy en día proliferan bajo las más variadas formas: animación por ordenador para la televisión y el cine, sistemas que reproducen objetos tridimensionales en perspectiva para arquitectos, decoradores, videojuegos y muchas cosas más. Los ingenieros denominan a las versiones que ellos utilizan sistemas de CAD (por Computer-Aided Design; Diseño Asistido por Ordenador). Los sistemas de CAD están revolucionando el mundo de la ingeniería, no sólo porque facili-

1. Esta útil simplificación es uno de los muchos detalles que posteriormente fueron objeto de estudio por parte de otros investigadores. Hoy se dispone de suficientes datos que evidencian la presencia de efectos de «inercia» y de «momento» en la transformación de imágenes. Véase Freyd (1989).

tan mucho el proceso de creación del diseño de la misma manera que los procesadores de textos facilitan mucho el proceso de redactar un documento, sino también porque con ellos los ingenieros pueden resolver problemas y responder preguntas que en otras circunstancias resultarían muy complejos. Enfrentado al problema de Shepard en la figura 10.1, un ingeniero podría responder a la pregunta con la ayuda de un sistema de CAD situando ambas imágenes en el TRC, haciendo rotar, literalmente, una de las imágenes e intentando superponer una de las imágenes sobre la otra. Hay algunos detalles de este proceso que nos interesan.

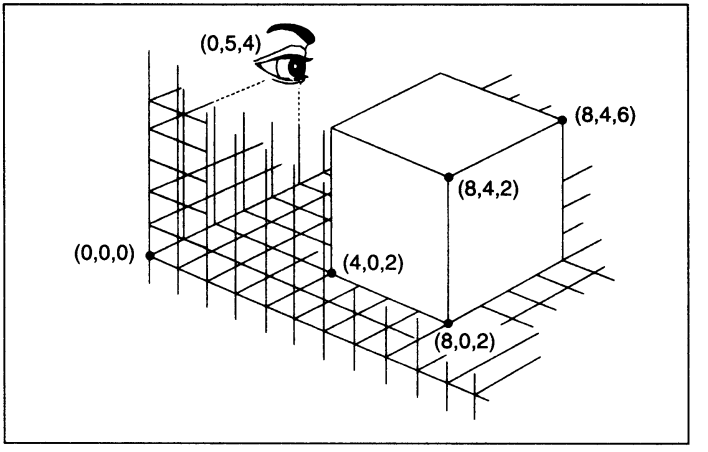
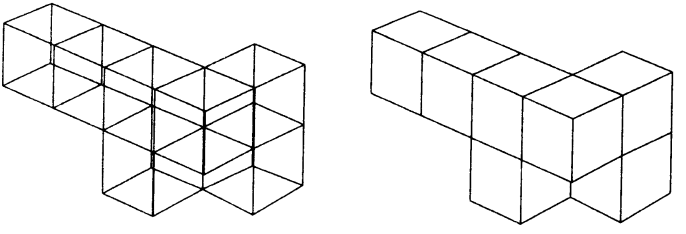


Figura 10.2

Cada uno de los objetos dibujados se introduciría en la memoria del ordenador como un objeto tridimensional *virtual*, descomponiéndolo en una descripción de sus planos y sus aristas definidos por sus coordenadas *xyz*, de modo que cada punto en el espacio virtual es un «triple ordenado» de números almacenado en la memoria del ordenador. El punto de vista de un observador implícito también se introduciría como un punto en el mismo espacio virtual, definido por su propio triple de coordenadas *xyz*. En la figura 10.2 tenemos el diagrama de un cubo y de un punto de vista, pero es importante no olvidar que lo único que el ordenador debe almacenar son los triples de cada punto clave, agrupados formando grupos mayores (por ejemplo, uno por cada cara del cubo), junto a información codificada de las distintas propiedades de cada cara (su color, si es opaca o transparente, su textura, etc.). La rotación de uno de esos objetos y su posterior desplazamiento en el espacio virtual son fáciles de calcular: basta con ajustar todas las coordenadas *x*, *y* y *z* en cantidades constantes; simple aritmética. Después, el calcular las líneas visibles que determinan qué planos del objeto serán visibles

desde el punto de vista virtual, y la forma exacta en que aparecerán, es también un simple problema de geometría. Los cálculos son simples pero laboriosos o de «computación intensiva», especialmente si también hay que calcular ligeras curvaturas, sombras, luz reflejada y texturas.

En los sistemas avanzados también es posible calcular los diversos fotogramas con la suficiente rapidez como para crear movimiento aparente en la pantalla, pero sólo si las representaciones se mantienen a un nivel suficientemente esquemático. La «supresión de líneas ocultas», el proceso por el cual se reproduce la imagen final de forma opaca en su situación correcta y que impide que un cubo de Shepard aparezca como un cubo de Necker transparente, es también un proceso que requiere su tiempo, lo cual fija aproximadamente los límites de lo que se puede producir «en tiempo real». Para obtener las maravillosas y detalladas transformaciones de imágenes realizadas con sistemas de gráficos por ordenador que vemos cada día en la televisión, el proceso de generación de imágenes es aún más lento, incluso en un superordenador, de modo que cada fotograma tiene que almacenarse para ser reproducido con posterioridad a mayor velocidad, a fin de satisfacer los requisitos que para la detección del movimiento impone el sistema visual humano.[2]



Antes y después de suprimir las líneas ocultas.

Figura 10.3

Estos manipuladores de objetos virtuales tridimensionales son unas excelentes nuevas herramientas o juguetes, y son realmente una novedad, no una copia electrónica de algo que ya tenemos en nuestras cabezas. No cabe duda de que ningún proceso análogo a estos miles de millones de cálculos geométricos y aritméticos se lleva a cabo en nuestros cerebros cuando producimos imágenes mentales, ni, por los motivos que ya examinamos en el capítulo 1, ninguna otra cosa *podría* reproducir las detalladas secuencias animadas a las que éstos procesos dan lugar.

2. La impresionante, aunque un tanto desigual, animación del popular programa «Flight Simulator» para IBM-PC, es una buena muestra de los límites de la animación en tiempo real en pequeños ordenadores de escenas tridimensionales bastante complejas.

Podemos quedar satisfechos con la idea de que esta limitación de nuestros cerebros es real, si consideramos un problema bastante parecido a los de Shepard, que sería muy fácil de resolver con la ayuda de un sistema de CAD: ¿Sería visible la X «roja» pintada sobre una de las caras de este objeto para alguien que mirara a través del agujero cuadrado abierto en la pared frontal?

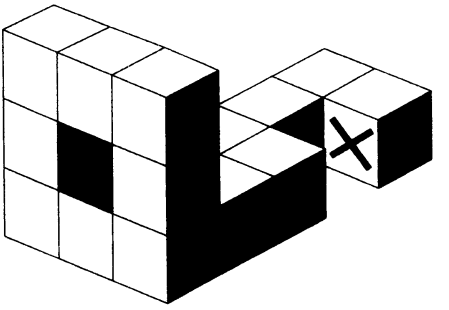


Figura 10.4

Nuestro objeto de Shepard con la X es un objeto simple y esquemático, y, puesto que la pregunta que queremos resolver es independiente de la textura, la iluminación y otros detalles estéticos, para un ingeniero sería bastante fácil producir una rotación animada de este objeto en el TRC. Podría, pues, hacer rotar la imagen en una u otra dirección, desplazando el punto de vista hacia adelante y hacia atrás, y simplemente buscar *vislumbrar* un poco de rojo a través del agujero. Si ve rojo, la respuesta es sí; en caso contrario, la respuesta es no.

Ahora bien, ¿puede usted llevar a cabo el mismo experimento en su ojo de la mente? ¿Puede usted hacer rotar el objeto y mirar a través del agujero? Si usted puede, entonces puede hacer algo que ni yo ni todos mis conocidos a los que he preguntado podemos hacer con un mínimo de fiabilidad. Incluso aquellos que tienen una respuesta para la pregunta están convencidos de que no sólo llegaron a ella haciendo la rotación y mirando. (A menudo cuentan que primero intentaron hacer la rotación, despues intentaron mirar, y vieron que no funcionaba; pueden «hacerla rotar» pero «se desmorona» cuando intentan mirar a través del agujero. Después hablan de intentar «dibujar» líneas de visión que cruzan el agujero de la imagen sin rotar, a fin de comprobar si pueden decidir si las líneas alcanzan el plano posterior.) Dado que nuestro objeto de Shepard no es más complejo que los objetos que los sujetos hacían rotar, aparentemente con éxito, en muchos experimentos, esto nos plantea un problema: ¿Qué tipo de proceso puede efectuar tan fácilmente algunas transformaciones (y después extraer información del re-

sultado), y fracasar tan estrepitosamente en otras operaciones que, en apariencia, no presentan muchas más exigencias? (Si estas operaciones no nos *parecen* mucho más complejas, debemos de estar examinándolas desde el punto de vista equivocado, ya que nuestros fracasos demuestran que sí son más exigentes.)

Un experimento llevado a cabo por los psicólogos Daniel Reisberg y Deborah Chambers (en preparación) plantea una cuestión muy parecida. Se mostraba a sujetos que afirmaban ser buenos figuradores unas formas «sin sentido», se les pedía que las hicieran rotar mentalmente 90 o 180 grados en los ojos de su mente y que relataran lo que «veían». Los sujetos se sorprendían al no poder reconocer, con los ojos de su mente, lo que usted podrá reconocer cuando gire el libro 90 grados en el sentido de las agujas del reloj y mire estas figuras.

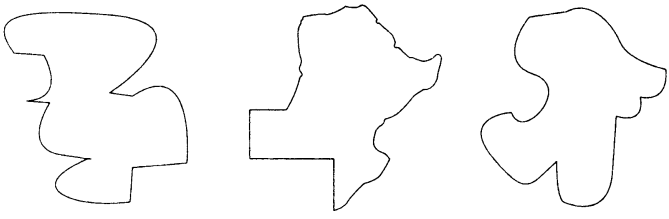


Figura 10.5

El tipo de preguntas que los ingenieros responden utilizando sistemas de CAD no suelen ser tan simples como «¿es visible la X roja a través del agujero?». Normalmente sus intereses se dirigen hacia propiedades espaciales más complicadas de los objetos que diseñan tales como «¿será capaz este brazo de robot con tres articulaciones de alcanzar el tornillo en la espalda del robot y ajustarlo sin chocar con la caja de la batería?» o incluso propiedades estéticas de estos objetos: «¿cómo quedará esta escalera en el vestíbulo del hotel, vista desde la calle a través de las cristaleras?». Cuando intentamos visualizar estas escenas sin ayuda, obtenemos unos resultados bastante burdos y poco fiables, de modo que un sistema de CAD podría entenderse como una especie de prótesis de la imaginación (Dennett, 1982d, 1990b). Es capaz de potenciar la capacidad de imaginación de los seres humanos, pero depende de que el usuario posea una capacidad de visión normal para mirar al TRC.

Imaginemos ahora un dispositivo protésico más ambicioso: ¡un sistema de CAD para ingenieros ciegos! Y para no complicar demasiado las cosas, supondremos que el tipo de preguntas que nuestros ingenieros ciegos quieren responder tienen relación con una geometría relativamente simple, y no preguntas sutiles sobre la estética de un diseño arquitectónico. Evidentemente, la salida del sistema deberá ser en un formato no visual. Quizás el forma-

to más práctico sea respuestas en lenguaje ordinario (en braille o como voz sintetizada) a preguntas en lenguaje ordinario. Así que vamos a suponer que cuando se enfrente a una pregunta del tipo que acabamos de considerar, el ingeniero ciego se limitará a transmitir la pregunta al sistema de CAD (en términos que éste pueda «comprender», por supuesto) y a esperar que el sistema de CAD responda.

Nuestro sistema CADBLIND/CADCIEGO Mark I no es muy elegante, pero es simple. Consta de un sistema de CAD ordinario, con su TRC, frente al cual se sienta un *Vorsetzer*, un sistema de visión por ordenador, con una cámara de TV enfocada al TRC y unos dedos de robot para accionar los botones del sistema de CAD.[3] A diferencia de Shakey, cuyo TRC estaba sólo para los espectadores, este sistema realmente «mira» la imagen, una imagen real hecha de puntos brillantes de fósforo, que irradia luz real a diferentes frecuencias sobre los traductores sensibles a la luz en la parte posterior de la cámara de TV. Cuando se le plantea nuestro problema de Shepard de la X roja, el CADBLIND Mark I produce una imagen con una X real, visible para todo el mundo, incluida la cámara de TV del Vorsetzer.

Supongamos, sin más, que el Vorsetzer ha resuelto en su interior los abundantes problemas de la visión por ordenador como para ser capaz de extraer la información buscada a partir de las representaciones que brillan en la pantalla del TRC. (No, no voy a afirmar que el Vorsetzer es consciente; simplemente quiero suponer que es lo bastante bueno haciendo lo que hace como para responder a las preguntas que el ingeniero ciego le pueda plantear.) El CADBLIND Mark I produce y manipula imágenes reales, y las utiliza para responder al ingeniero ciego todas las preguntas que un ingeniero con vista normal podría responder utilizando un sistema de CAD ordinario. Si el sistema Mark I es bueno, entonces será un juego de niños diseñar el Mark II: nos deshacemos del TRC y de la cámara de TV enfocada hacia él y los sustituimos por un simple cable. A través de este cable, el sistema de CAD envía al Vorsetzer el *mapa de bits*, la matriz de ceros y unos que define la imagen en el TRC. En el Vorsetzer del Mark I, este mapa de *bits* se reconstruía trabajosamente a partir de la información de salida enviada por los traductores ópticos de la cámara.

En cuanto a *computación*, apenas nos ahorramos algo con el Mark I; sólo hemos eliminado algún *hardware* innecesario. Todos esos cálculos tan complejos para las líneas de visión, para eliminar líneas ocultas y para la reproducción de texturas, sombras y luz reflejada, que requerían tanto tiempo de computación en el Mark I, siguen formando parte del proceso. Supóngase que se solicita al Vorsetzer que emita un juicio sobre profundidades comparando gradientes de texturas o interpretando unas sombras. Tendrá que *ana-*

3. Denomino *Vorsetzer* a este dispositivo porque me recuerda a un maravilloso pianista-robot alemán que también se llamaba así, que constaba de una unidad separada con ochenta y ocho «dedos» y que se podía «sentar ante» un piano ordinario, presionando las teclas y los pedales desde el exterior, igual que un pianista humano. (Es importante caer en la cuenta de que este dispositivo es un *Vorsetzer* —uno que se sienta ante algo— pero no un *Vorsitzer* —un presidente.)

lizar los patrones de *bits* que aparecen en las áreas relevantes del mapa de *bits*, a fin de llegar a discriminar las texturas y las sombras.

Ello significa que el Mark II es todavía una máquina de una gran ineficiencia, ya que si la información de que una determinada zona del mapa de *bits* debiera representar una sombra, eso es algo que el sistema de CAD ya «sabe» (si ello es parte de la descripción codificada del objeto a partir de la cual el sistema de CAD genera sus imágenes), y si este hecho es parte de lo que el Vorsetzer debe determinar a fin de emitir su juicio, ¿por qué el sistema de CAD no se lo *dice* al Vorsetzer? ¿Por qué molestarse en *reproducir* la sombra en beneficio de los analizadores de patrones del Vorsetzer, si la tarea de reproducción de patrones y la tarea de análisis de patrones se anulan mutuamente?

Así pues, nuestro CADBLIND Mark III se verá libre de todas esas tareas computacionales tan pesadas de reproducción de imágenes: toma la mayor parte de lo que ya «sabe» sobre los objetos representados y se lo envía directamente al subsistema Vorsetzer, utilizando el formato de simples códigos para propiedades y prendiendo «etiquetas» sobre las diferentes «zonas» de la matriz del mapa de *bits*, que, de este modo, pasa de ser una pura imagen a ser algo parecido a un diagrama. *Algunas* propiedades espaciales se representan directamente —*se muestran*— en el espacio (virtual) del mapa de *bits*, pero otras simplemente *se refieren* mediante etiquetas.[4]

Esto debería recordarnos la afirmación que hice en el capítulo 5 de que el cerebro tiene que efectuar sus discriminaciones sólo una vez; un rasgo que ha sido identificado no tiene por qué volver a presentarse en beneficio del maestro apreciador en el Teatro Cartesiano.

Ahora estamos en disposición de apreciar un nuevo aspecto de la ingeniería: la «anulación» sólo funciona si los sistemas que necesitan comunicarse «hablan el mismo lenguaje». ¿Qué ocurriría si el formato con que el sistema de CAD ya «conoce» la información relevante —por ejemplo, la información de que algo es una sombra— no es un formato que permita al Vorsetzer «utilizar» esa información?[5] Así pues, a fin de que la comunicación tenga éxito puede ser necesario «dar un paso atrás para poder saltar hacia adelante». Podría ser necesario que los sistemas se enzarzasen en una serie de interacciones disolutas —diríase que interminables— desde el punto de vista de la información a fin de poder interactuar realmente. Piense usted en la situación de dibujar un mapa para dar indicaciones a un extranjero, cuando todo lo que tiene que saber —ojalá pudiera decírselo en su propia lengua— es «tuerza a la izquierda en el próximo semáforo». A veces, tomarse todas las molestias que comporta hacer una imagen es necesario desde el punto de vista práctico, incluso si no es necesario «en principio».

4. Una vez tenemos etiquetas, podemos explicar *cualquier* propiedad de un objeto, no sólo propiedades espaciales o propiedades visibles; como en los viejos chistes de los cuadernos para colorear: «Ahí está mi jefe, píntalo bien feo».

5. Véase Kosslyn (1980) para algunas observaciones sobre los formatos. Jackendoff (1989) tiene un análisis parecido de lo que él denomina la forma de las estructuras de información.

Habida cuenta de que los sistemas en nuestros cerebros son el producto de muchas historias de oportunos remiendos superpuestos, la larga historia de la selección natural y la corta historia del rediseño individual por automanipulación, no es de extrañar que existan tales ineficiencias. Además, hay otros motivos para reproducir la información con un formato figurativo (a parte del mero placer de hacerlo), que, en caso de que demos con ello por casualidad, pronto nos harán tener la impresión de que hacer imágenes siempre vale la pena. Como ya señalamos en nuestras especulaciones del capítulo 7 sobre «dibujar para uno mismo», dichas transformaciones de formato constituyen a menudo modos muy efectivos para extraer información que en caso contrario resulta impenetrable a partir de los datos. Los diagramas sí son re-*presentaciones* de la información, pero no para un ojo interno, sino para un mecanismo de reconocimiento de patrones interno que también puede recibir información de un ojo (¿«externo»?) ordinario. Éste es el motivo por el cual las técnicas de gráficos (por ordenador) son tan útiles para la ciencia, por ejemplo. Permiten que grandes cantidades de datos se presenten en un formato que facilite que las excelentes capacidades para el reconocimiento de patrones de la visión humana se hagan cargo de todo. Hacemos gráficos y mapas y todo tipo de dibujos coloreados a fin de que las regularidades y los elementos más notables que buscamos «se aparezcan» ante nosotros, gracias a nuestros sistemas visuales. Los diagramas no sólo nos ayudan a percibir regularidades que en caso contrario serían imperceptibles; nos permiten *hacer un seguimiento* de lo que es relevante, y nos *recuerdan* que hagamos las preguntas adecuadas en los momentos precisos. El investigador sueco en IA Lars-Erik Janlert (1985) ha argumentado que dichos actos de generación y de examen de imágenes en un ordenador pueden utilizarse también como ayuda para resolver problemas que de otro modo serían intratables en lo que podríamos denominar gestión de inferencias en sistemas que «en principio» son meros motores deductivos. (Para algunas observaciones sobre el mismo proceso con un sesgo ligeramente distinto, véase Larkin y Simon, 1987.)

Esta estrategia es bien conocida por muchos hábiles pensadores y ha sido objeto de una excelente descripción por parte de uno de los más hábiles, el físico Richard Feynmann, en *Surely You're Joking, Mr. Feynmann!* (1985). En un capítulo oportunamente titulado «Otra caja de herramientas», nos cuenta cómo dejaba asombrados a sus condiscípulos de Princeton «intuyendo» la verdad o falsedad de ciertos oscuros teoremas de topología, teoremas que era completamente incapaz de derivar formalmente o incluso llegar a comprender por completo:

> Tenía una técnica, que aún hoy utilizo cuando alguien está explicándome algo que estoy intentando comprender: no dejo de construir ejemplos. Por ejemplo, los matemáticos aparecen con un teorema fabuloso y están todos muy animados. A medida que me van exponiendo las condiciones del teorema, yo construyo algo que cumple todas las condiciones. Ya se sabe, tenemos un conjunto

(una bola) —disjunto (dos bolas)—. Entonces, las bolas toman color, les crece el pelo o cualquier otra cosa en mi cabeza a medida que se van añadiendo condiciones. Finalmente enuncian el teorema, que resulta ser alguna tontería sobre la bola y que resulta no ser verdad de mi bola con el pelo verde, así que digo, «¡falso!».

Si es verdad, todos se acaloran mucho y yo les dejo continuar un rato más. Finalmente, les pongo mi contraejemplo.

«Ah, bueno. Es que nos hemos olvidado de decirte que es un homomorfismo de Hausdorff de Clase 2.»

«Vale», respondo, «¡entonces es trivial, es trivial!» En ese momento ya sé cómo funciona, aunque no sepa qué es un homomorfismo de Hausdorff (págs. 85-86).

Estas estrategias, hasta cierto punto, «surgen naturalmente», pero hay que aprenderlas o inventarlas, y algunas personas lo hacen mucho mejor que otras. Aquellas personas en que estas destrezas están más desarrolladas poseen máquinas virtuales diferentes en sus cerebros, con poderes significativamente distintos, comparadas con las de aquellos que son «visualizadores» ineptos o poco comunes; y estas diferencias surgen inmediatamente en sus mundos heterofenomenológicos.

Así pues, hay buenas razones para creer, como Kosslyn y otros han propuesto, que los seres humanos ponen a trabajar sus sistemas visuales no sólo presentándose imágenes reales *externas* (como en el TRC de un sistema de CAD), sino también con imágenes virtuales internas diseñadas para la ocasión o con representaciones de datos en forma de diagrama que son la materia prima apropiada para el estadio o estadios posteriores de los mecanismos de procesamiento de la visión.

¿Qué soluciones de ingeniería para qué problemas de comunicación interna y manipulación de la información ha hallado el cerebro humano, y cuáles son sus puntos fuertes y sus puntos flacos?[6] Éstas son las preguntas empíricas que se plantea la investigación en el ámbito de la psicología cognitiva, y deberíamos ser muy cautos en el momento de avanzar respuestas *a priori*.[7] Supongo que podríamos haber encontrado los sistemas de manipulación de imágenes del Mark I en nuestros cerebros, al completo, con sus puntos de fósforo brillante y su ojo interno sensible a la luz. (Por lo que yo puedo ver, no sería *imposible* que las criaturas de algún planeta estuvieran dotadas con dichos dispositivos.) Y se necesitan experimentos como los de Reisberg y Chambers para demostrar que los atajos que encuentran nuestros ce-

6. Kosslyn (1980) no sólo presenta una defensa detallada de sus propias respuestas para estas preguntas, sino que también presenta un detallado examen de otros trabajos experimentales y teóricos sobre la imaginería. Farah (1988), y Finke, Pinker y Farah (1989) constituyen dos buenos repasos de los trabajos realizados durante la siguiente década.

7. En palabras de Marvin Minsky, «La idea de percibir sucesos internos del cerebro no tiene nada de singular. Los agentes son agentes, y es tan sencillo para uno de ellos estar conectado para detectar un *suceso cerebral causado por el cerebro*, como estarlo para captar un *suceso cerebral causado por el mundo*» (1985, pág. 151) [pág. 156 de la traducción castellana; N. del T.].

rebros casi confirman la imposibilidad de descubrir un sistema Mark II, con un formato de mapa de *bits* que nunca aprovecha las ventajas de tomar un atajo. (Si dispusiéramos de un sistema como éste, resolver el problema de la X roja sería fácil, y también lo sería el de hacer rotar el Estado de Texas.)

La fenomenología proporciona algunas pistas que apuntan en ambas direcciones: el «carácter burdo» de las imágenes mentales, que es «intuitivamente obvio» en la fenomenología de la mayor parte de sujetos, apunta hacia el uso de atajos por parte del cerebro, ocasiones en las que el cerebro cuenta sin mostrar. Esto vale tanto para la *percepción* visual como para la *visualización*. Ya comentamos en el capítulo 2 lo difícil que resulta el dibujar una rosa que tenemos ante nuestros ojos, e incluso copiar un dibujo, y el motivo de estas dificultades reside en el hecho de que las propiedades eminentemente espaciales que uno debe identificar y discriminar a fin de dibujar bien normalmente se quedan atrás en el curso del procesamiento perceptivo, resumido en testimonios, no reproducidos para exámenes posteriores. Por otra parte, la utilidad de las imágenes mentales para ayudarnos a «percibir un patrón» o para «recordarnos» detalles que podríamos olvidar, apunta hacia una explotación de los mecanismos visuales de reconocimiento de patrones que solamente podría producirse si una parte del cerebro se tomara la molestia de preparar versiones de la información en un formato especial para dichos sistemas visuales. Como ya vimos en el capítulo 1, las exigencias en cuanto a tareas de tratamiento de la información para dicha representación son formidables, y no debe sorprendernos que seamos tan poco hábiles en mantener estables en nuestras mentes incluso diagramas muy esquemáticos.

He aquí una simple prueba para recordarle cuán limitadas son nuestras capacidades: *en su ojo de la mente*, rellene el siguiente crucigrama de tres por tres, escribiendo las siguientes tres palabras en las *verticales*, empezando por la columna de la izquierda: GEN, ALA, SEO.

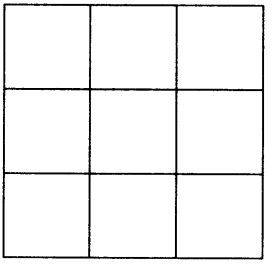


Figura 10.6

¿Es usted capaz de *leer* las palabras horizontales? En un diagrama sobre papel, estas palabras se le «aparecerían»; sería usted incapaz de *no* verlas. Ésta es, precisamente, la utilidad de hacer diagramas: presentar los datos con un formato que facilita o hace inevitable un desglose o un análisis novedosos de dichos datos. Un damero de tres por tres con caracteres alfabéticos no es un estructura de datos muy complicada, aunque, aparentemente, es algo que nuestros cerebros no pueden mantener de forma bastante estable para que sus sistemas visuales efectúen su tarea de «hacerlo aparecer». (Si usted quiere volverlo a intentar, aquí tiene dos grupos de palabras más para las verticales: POR, ISA, MÁS y ERA, SIN ASA.)

Sin embargo, hay un amplio abanico de posibilidades para la variación individual de las estrategias empleadas por los diferentes visualizadores, y muchos pueden ser capaces de hallar —o desarrollar— estrategias de figuración que les permitan «leer» estos diagramas. Los prodigios del cálculo mental pueden instruirse a sí mismos en la multiplicación de números de diez dígitos, de modo que no sería sorprendente que algunas personas pudieran desarrollar un talento para la «lectura de dameros» en sus ojos de la mente. Estas pruebas informales nos dan algunas pistas, pero son los experimentos los que pueden definir de forma más clara los tipos de mecanismos y procesos que las personas deben de estar usando en estos actos de automanipulación. Por el momento, las evidencias apoyan una visión según la cual utilizamos una estrategia mixta, sacando partido del análisis de matrices, pero también incorporando el atajo del etiquetado, *referir* sin *mostrar*.

Nótese, sin embargo, que incluso en el sistema CADBLIND Mark II, que es ultrafigurativo y que incorpora un mapa de *bits* que reproduce el color, las sombras y las texturas *píxel* a *píxel*, hay un sentido —un sentido metafísicamente importante, además, como podremos comprobar en los dos capítulos siguientes— en el que todo es «referir» y nada es «mostrar». Considérese la X roja sobre nuestra figura de Shepard (figura 10.4). En el Mark I, ésta se reproduce en rojo real: el TRC emite luz, que debe ser traducida por algo en la cámara de TV parecido a los conos en nuestros ojos que responden a las diferencias de frecuencia. Cuando el Vorsetzer hace rotar la imagen hacia adelante y hacia atrás, con el objetivo de vislumbrar un poco de rojo a través del agujero, está esperando que sus demonios detectores de rojo empiecen a chillar. En el Mark II, eliminamos todo este *hardware*, y el mapa de *bits* representa el color de cada *píxel* mediante números. Supongamos que el tono de rojo es el número 37. El Vorsetzer en el Mark II, cuando hace rotar la imagen codificada en un mapa de *bits* hacia adelante y hacia atrás, está mirando a través del agujero esperando vislumbrar el 37. O, en otras palabras, está preguntando a los demonios-*píxel* si quieren decirle, «color número 37 aquí». El rojo ya no está, sólo quedan números. En última instancia, todo el trabajo en un sistema CADBLIND debe hacerse mediante operaciones aritméticas entre secuencias de *bits*, como vimos en el capítulo 4 que ocurre en el nivel más bajo de Shakey. Y no importa lo cuasipictóricos o figurativos que sean los procesos que resultan en las respuestas verbales

del Vorsetzer; éstos no se generarán en un lugar interno donde las propiedades perdidas (las propiedades que sólo «se refieren» en el mapa de *bits*) se ven *restituidas* a fin de ser *apreciadas* por un juez que compone las respuestas.

Las personas no son sistemas CADBLIND. El hecho de que un sistema CADBLIND pueda manipular e inspeccionar sus «imágenes mentales» sin la participación de un Teatro Cartesiano no prueba en sí mismo que no haya Teatro Cartesiano en el cerebro humano, pero prueba que no necesitamos postular un Teatro Cartesiano para explicar el talento humano de resolver problemas «en el ojo de la mente». Efectivamente, existen procesos que son muy parecidos a la observación, pero cuando despojamos la metáfora del TRC de Kosslyn de todo lo que resulta superfluo, eliminamos precisamente aquellos rasgos que hacen necesario el Teatro Cartesiano. No tiene por qué haber un tiempo y un espacio donde «todo se junta» en beneficio de un único y unificado discriminador; las discriminaciones pueden llevarse a cabo de modo distribuido, no sincrónico y multinivel.

2. Palabras, imágenes y pensamientos

> El aspecto verdaderamente «creativo» del lenguaje no reside en su «capacidad generativa infinita», sino en ciclos de producción y comprensión mediados por una mente capaz de reflexionar sobre los múltiples significados asociables a una proferencia, significados que no tienen por qué estar presentes en el pensamiento que dio lugar a la proferencia, pero que están disponibles gracias a la autocomprensión (o la interpretación profunda de la proferencia de otro) y que pueden dar lugar a un nuevo pensamiento que será expresado y reinterpretado y, así, sucesivamente, hasta el infinito.
>
> H. STEPHEN STRAIGHT, 1976, pág. 540

Al economista británico John Maynard Keynes le preguntaron una vez si pensaba con palabras o con imágenes. «Yo pienso con pensamientos», contestó él. Considero acertada su reacción de resistirse a la idea de que «las cosas con que pensamos» son o palabras o imágenes, ya que, como hemos podido comprobar, «las imágenes mentales» no son *únicamente* imágenes en la cabeza, ni el pensamiento «verbal» consiste *únicamente* en hablar con uno mismo. Sin embargo, decir que uno piensa con pensamientos no es mucho mejor respuesta. En todo caso, no hace más que posponer el asunto, ya que un pensamiento es precisamente *aquello que se produce cuando pensamos*, y no es ésta una cuestión sobre la que exista consenso.

Ahora que hemos considerado algunos esbozos del tipo de mecanismos subyacentes que son causalmente responsables de los detalles de nuestros mundos heterofenomenológicos, podemos *empezar* a dar cuenta de la fenomenología del pensamiento, explicando no sólo las limitaciones y los condicionamientos de la fenomenología «visual» y «verbal», sino también tratando de hallar otras variedades que escapen a esta dicotomía.

Uno de mis ejercicios preferidos de heterofenomenología novelada es la novela de Vladimir Nabokov *La defensa* (1930),* sobre el gran maestro Luzhin, un genio del ajedrez, que sufre una depresión nerviosa justo en mitad de la más intensa de sus partidas. Presenciamos tres estadios en el desarrollo de su conciencia: su mente infantil (antes de su descubrimiento del ajedrez a la edad de diez años), su mente saturada de ajedrez (hasta el momento de la depresión nerviosa), y los penosos restos de las dos etapas anteriores después de la depresión, cuando, recluido por su esposa en un mundo sin ajedrez —sin hablar de ajedrez, sin jugar al ajedrez, sin libros de ajedrez—, su mente sufre una regresión hacia una especie de paranoia de niño consentido, iluminada por perdidos momentos de ajedrez —fugaces y furtivos ataques en los problemas de ajedrez del periódico—, pero que acaba por cuajarse en obsesiones de ajedrez que culminan en su «jaque mate». Luzhin, nos cuenta la novela, tiene la mente tan saturada de ajedrez que percibe toda su vida en estos términos. En estos breves pasajes presenciamos la torpe manera que tiene de cortejar a la mujer que será su esposa:

> Luzhin empezó con una serie de movimientos prudentes, cuyo significado apenas podía percibir él mismo, su peculiar declaración de amor. «Vamos, sigue, cuéntame más», repetía ella, a pesar de haber percibido su semblante silencioso y taciturno.
>
> Él permaneció así, sentado, apoyándose en su bastón, y pensando que, con un movimiento de caballo de ese tilo que crecía en la colina bañado por el sol, uno podría comerse aquel poste telegráfico, y, a la vez, estaba intentando recordar qué había empezado a decir exactamente (pág. 97).*

> Con su hombro recostado sobre su pecho, ella intentó, con un dedo prudente, levantarle un poco más los párpados, y, con la ligera presión que ejerció sobre su globo ocular, se produjo un extraño brillo negro, saltando como su caballo negro que se comería al peón si Turati lo hiciera avanzar en el séptimo movimiento, como así ocurrió en su último encuentro (pág. 114).

He aquí una muestra de su estado después de la depresión:

> Se encontró en un local lleno de humo y de ruidosos fantasmas sentados ante sus mesas. En cada rincón se estaba desarrollando un ataque —y, apartando mesas, un cubo del que emergía un peón de cristal con el cuello dorado y un tambor que tocaba un caballo encabritado de espesas crines, dirigió sus pasos hacia una puerta giratoria que giraba lentamente... (pág. 139).

En muchos sentidos estos temas son «imágenes», ya que el ajedrez es un juego espacial, en el que incluso la identidad de las piezas se define por sus formas, pero el poder del ajedrez sobre la mente de Luzhin no se agota con

* No tengo noticia de que esta obra haya sido traducida al castellano. Existe, sin embargo, una traducción francesa, *La défense Loujine*, publicada por Gallimard/Folio, y una traducción inglesa (la citada por Dennett), *The Defense*, edición de G. P. Putnam, publicada por Popular Library en 1964 (aparecida originalmente en *The New Yorker*). [N. del T.]

* Las páginas remiten a la edición inglesa de 1964. [N. del T.]

las propiedades visuales o espaciales del ajedrez: cualquier cosa que pudiera ser capturada en fotografías o películas del tablero, sus piezas en movimiento. Sin embargo, estas propiedades visuales no le dan más que un sabor superficial a su imaginación. Mucho más poderosa es la *disciplina* de las reglas y la táctica del juego; es con la estructura abstracta del ajedrez con la que ha llegado a familiarizarse de forma tan obsesiva, y con sus hábitos de exploración de esta estructura con los que lleva a su mente de «pensamiento» a «pensamiento».

> Luzhin, con gran desesperación, pronto caería en la cuenta de que había vuelto a comportarse como un incauto y que en su vida se había vuelto a llevar a cabo un movimiento delicado, continuando despiadadamente la combinación fatal. Decidió, entonces, que debía intensificar la vigilancia y seguir con atención cada segundo de su vida, pues puede haber trampas por todas partes. Y se sentía oprimido principalmente por su incapacidad de inventar una defensa racional, ya que el objetivo de su oponente permanecía oculto (pág. 227).

Cuando usted aprendió a ir en bicicleta o a conducir, se encontró con una nueva estructura de oportunidades para la acción, con restricciones, señales, carriles, vistas, una especie de laberinto comportamental por el que usted ha aprendido a moverse rápidamente. Pronto se convirtió en una «segunda naturaleza». Usted pronto incorporó la estructura de este fenómeno externo en su propia estructura de control. Durante el proceso, puede que haya pasado por períodos de exploración obsesiva, cuando no podía apartar sus pensamientos de las nuevas reglas. Recuerdo un breve período de bridgemanía durante mi adolescencia, durante el cual tuve sueños obsesivos y sin sentido sobre el bridge. Hacía el mismo impase más de cien veces, o soñaba con el «remate» durante las conversaciones con mis profesores y compañeros de clase. Mis ensoñaciones hipnogógicas (esos períodos de alucinación por los que pasan ocasionalmente las personas cuando están empezando a dormirse o cuando están empezando a despertarse) estaban llenas de problemas como por ejemplo: «Cuál es la respuesta correcta a un remate de tres libros, ¿cuatro cuchillos o cuatro tenedores?».

Cuando uno encuentra una nueva estructura abstracta en el mundo —la notación musical, un lenguaje de programación, el código civil, la liga de béisbol— es bastante habitual que se encuentre recorriendo todos sus caminos y creando nuevas rutas mentales por uno mismo; trillando el campo para poder sentirse como en su casa. El de Luzhin no es más que un caso extremo; sólo tiene una estructura con la que jugar, y la utiliza para todo. Finalmente, su estructura domina todas las demás estructuras de hábitos de su mente, canalizando la secuencia de sus pensamientos con casi tanta rigidez como la secuencia de instrucciones de un programa en una máquina de von Neumann.

Piense en todas las estructuras que ha aprendido, en la escuela o en cualquier parte: decir la hora, la aritmética, el dinero, las líneas de autobús, usar

el teléfono. Pero de todas las estructuras que llegamos a aprender en el curso de nuestras vidas, no cabe duda de que la fuente más omnipresente y poderosa de disciplina en nuestras mentes es nuestra lengua materna. (A veces, las cosas se comprenden mejor por contraste; Oliver Sacks, en *Veo una voz*, 1989, nos hace ver con gran claridad la riqueza que el lenguaje aporta a la mente, refiriéndonos el terrible empobrecimiento que sufre la mente de un niño sordo, si a este niño se le niega todo acceso a una lengua *natural* como la seña o cualquier otro lenguaje de signos.) En el capítulo 8 vimos de qué manera el vocabulario a nuestra disposición influye no sólo en el modo en que hablamos con los demás, sino también en el modo en que hablamos con nosotros mismos. Pero por encima de la contribución *léxica* está la contribución *gramatical*. Como señala Levelt (1989, §3.6), las estructuras obligatorias de las oraciones de nuestras lenguas son como unos guías, recordándonos que verifiquemos esto, que prestemos atención a aquello, *exigiéndonos* que organicemos las cosas de una determinada manera. Sin duda parte de esta estructura puede ser innata, según han argumentado Chomsky y otros, aunque no constituye un problema fundamental el llegar a dilucidar dónde está la línea divisoria entre las estructuras que han sido depositadas genéticamente en el cerebro y aquellas que han entrado en forma de memas. Dichas estructuras, reales o virtuales, tienden algunas de las vías por donde circularán los «pensamientos».

El lenguaje infecta y modula nuestro pensamiento a todos los niveles. Las palabras de nuestro vocabulario son catalizadores que pueden hacer precipitar fijaciones de contenido mientras una parte del cerebro intenta comunicarse con otra. Las estructuras de la gramática inculcan disciplina a nuestros hábitos de pensamiento, conformando las maneras en que sondeamos nuestras «bases de datos», intentando, como el criador de pájaros de Platón, conseguir que acudan los pájaros adecuados cuando los llamamos. La estructura de las historias que aprendemos sirve de guía a diferentes niveles, incitándonos a hacernos las preguntas con más posibilidades de ser relevantes en las circunstancias del momento.

Nada de esto tiene sentido mientras persistamos en pensar que la mente es de una racionalidad ideal, y de una transparencia para sí y una unificación perfectas. ¿Qué bien puede hacer el *hablar con uno mismo*, si ya se sabe lo que se quiere decir? Pero una vez vislumbramos una posibilidad de entendimiento, aunque sea parcial, de imperfecta racionalidad, de intercomunicación problemática entre las partes, podemos ver de qué modo pueden explotarse las poderosas fuerzas que libera un lenguaje en el cerebro mediante varias formas de autoayuda, algunas beneficiosas y algunas perjudiciales.

Éste es un ejemplo:

«¡Eres genial!».

Éste es otro:

«¡Eres patético!».

Usted sabe lo que significan estas frases. También sabe que las he introducido de repente, como ayuda para un argumento filosófico, y que no pretenden ser los actos de habla de nadie. En efecto, ni lo estoy halagando, ni lo estoy insultando, y aquí no hay nadie más. ¿Pero podría usted halagarse a sí mismo, o insultarse a sí mismo, tomando una de estas frases mías, y diciéndosela a sí mismo, una y otra vez, «con énfasis»? Inténtelo, si se atreve. Algo ocurre. Por un momento, usted no se cree a sí mismo (se dice usted), pero descubre que eso de decirse estas palabras provoca reacciones, quizás incluso un pequeño enrojecimiento de orejas, junto a otras respuestas, réplicas, rechazos, imágenes, recuerdos, proyectos. Dichas reacciones pueden moverse en una dirección o en otra, evidentemente. Dale Carnegie estaba en lo cierto sobre los poderes del pensamiento positivo, pero como la mayoría de las tecnologías, el pensamiento es más fácil de crear que de controlar. Cuando uno habla consigo mismo, no tiene por qué creerse para provocar reacciones. Tienen que ser reacciones y, de un modo u otro, tienen que ser relevantes para el significado de las palabras con las que uno se está estimulando. Una vez las reacciones empiezan a producirse, éstas pueden conducir a la mente hasta lugares donde uno acaba por creerse a sí mismo después de todo, así que vaya usted con cuidado con las cosas que se dice.

El filósofo Justin Leiber resume el papel que juega el lenguaje en la formación de nuestra vida mental:

> Mirándonos desde el punto de vista de un ordenador, no podemos evitar ver que el lenguaje natural es nuestro «lenguaje de programación» más importante. Ello significa que una enorme porción de nuestros conocimientos y actividades, para nosotros, debe ser comunicada y comprendida en nuestra lengua... Podría decirse que el lenguaje natural fue nuestro primer gran artefacto original y puesto que, como vamos percibiendo cada vez con mayor claridad, los lenguajes son máquinas, el lenguaje natural, con el cerebro sobre el que corre, fue nuestra primera invención del ordenador universal. Todo esto podría decirse, si no fuera porque permanece la sospecha de que el lenguaje no es algo que hayamos inventado sino algo en lo que nos hemos convertido, no algo que construimos sino algo con lo que nos creamos, y recreamos, a nosotros mismos (Leiber, 1991, pág. 8).

La hipótesis de que el lenguaje juega un papel fundamental en el pensamiento podría parecer a primera vista una versión de la largamente debatida hipótesis de que existe un *lenguaje del pensamiento*, un único medio dentro del cual se produce toda cognición (Fodor, 1975). Existe, no obstante, una importante diferencia. Con gran acierto, Leiber equipara el lenguaje con un lenguaje de programación para el cerebro, pero es preciso distinguir los lenguajes de programación de alto nivel (como el Lisp, el Prolog o el Pascal) del básico «lenguaje máquina» y el un poco menos básico «lenguaje ensamblador» a partir de los cuales se componen los primeros. Los lenguajes de

alto nivel son máquinas virtuales y crean estructuras (temporales) en un ordenador, que lo dotan de un determinado grupo de puntos fuertes y de puntos débiles. El precio que uno paga por poder hacer que ciertas cosas sean «fáciles de decir» es hacer que otras cosas sean «difíciles de decir» o incluso imposibles. Una máquina virtual como ésta puede tener la capacidad de estructurar sólo una parte de la competencia del ordenador, sin llegar a afectar otras partes de la maquinaria subyacente. Si tenemos en mente esta distinción, entonces es lícito mantener que los detalles de un lenguaje natural —el vocabulario y la gramática del inglés, el chino o el castellano— imponen constricciones sobre el cerebro de la misma manera que un lenguaje de programación de alto nivel. Pero esto es muy distinto de defender la dudosa hipótesis de que dicho lenguaje natural establece la misma estructura *a todos los niveles*. Evidentemente, Fodor y los que también defienden la idea de un lenguaje del pensamiento insisten en afirmar que *no* están hablando del nivel en el que los lenguajes *humanos* llevan a cabo su tarea de imposición de restricciones. Están hablando de un nivel de representación más profundo y menos accesible. En cierta ocasión, Fodor defendió su postura con la ayuda de una divertida confesión: confesó que cuando estaba pensando con todas sus fuerzas, los únicos elementos lingüísticos de los que era consciente eran pequeñas secuencias como «¡vamos, Jerry, que tú puedes!». Puede que éstos hayan sido sus «pensamientos», y acabamos de ver que pueden haber jugado un papel muy importante en ayudarle a resolver los problemas con que se enfrentaba, pero distan mucho de ser la sustancia con que se forjan las inferencias perceptivas, con que se verifican las hipótesis y con que se llevan a cabo las demás transacciones del lenguaje del pensamiento básico. Keynes estaba en lo cierto al resistirse a aceptar la dicotomía entre palabras e imágenes; los medios que utiliza el cerebro son sólo vagamente parecidos a los medios de representación de la vida pública.

3. Referir y expresar

Sin prisa, pero sin pausa, hemos ido minando la idea del Teatro Cartesiano. Esbozamos una alternativa al Significador Central en el capítulo 8, y hemos visto cómo podemos resistirnos al atractivo de un TRC interno. Muchos golpes mal dirigidos, me temo, pues el Teatro Cartesiano sigue ahí, ejerciendo aún su poder sobre nuestra imaginación. Ha llegado el momento de cambiar de táctica y de atacar desde dentro, haciendo estallar el Teatro Cartesiano demostrando su incoherencia en sus propios términos. Veamos lo que ocurre cuando, siguiendo la tradición, aceptamos la «psicología del sentido común» de todos los días en sentido literal. Podemos empezar reconsiderando algunas de las afirmaciones plausibles que hizo Otto al principio del capítulo 8:

> Cuando *yo* hablo, [dijo Otto] quiero decir lo que digo. Mi vida consciente es privada, pero puedo decidir hacerle a usted partícipe de algunos de sus aspectos. Puedo decidir contarle cosas sobre mi experiencia pasada o presente. Cuando lo hago, formulo enunciados que adapto cuidadosamente al material que deseo relatar. Puedo ir de la experiencia al testimonio propuesto contrastando las palabras con la experiencia para asegurar que he dado con *les mots justes*... Atiendo a los dictados de mi experiencia consciente y llego así a un juicio sobre las palabras que mejor le harán justicia a su carácter. Cuando me siento satisfecho de haber elaborado un testimonio preciso, lo expreso. A partir de mi relato introspectivo, usted puede llegar a conocer alguna característica de mi experiencia consciente.

Parte de este mensaje encaja bastante bien en el modelo de producción del lenguaje que propusimos en el capítulo 8. Ese proceso de ir y venir, adecuando las palabras al contenido de la experiencia podemos verlo también en el pandemónium que empareja los demonios-palabra con los demonios-contenido. Lo que falta, por supuesto, es el Yo Interno cuyos juicios dirigen ese emparejamiento. Pero aunque Otto sigue hablando de lo que «yo elijo» y de lo que «yo juzgo», la introspección no confirma esta visión.

Tenemos un acceso muy limitado a los procesos por los cuales «se nos ocurren» las palabras que diremos, incluso en aquellos casos en que hablamos deliberadamente, ensayando nuestros actos de habla en silencio antes de proferirlos. Los candidatos a ser algo que decir aparecen simplemente de nadie sabe dónde. O bien ya nos encontramos diciéndolos, o bien nos encontramos verificándolos, a veces descartándolos, otras corrigiéndolos un poco y después diciéndolos, pero ni siquiera estos ocasionales pasos intermedios nos ofrecen nuevas pistas sobre cómo los hacemos. Simplemente nos encontramos descartando esta palabra o aquella. Si tenemos alguna motivación para nuestros juicios, rara vez los contemplamos antes del acto, sino que sólo nos parecen obvios en retrospectiva. («Iba a utilizar la palabra *infecundo* pero me detuve, porque me sonaba demasiado pretenciosa.») Así pues, no poseemos ninguna capacidad de penetración en los procesos que se producen en nuestro interior durante el paso del pensamiento al habla. Por lo que sabemos, *podrían* perfectamente ser el producto de un pandemónium.

> Y sin embargo, [prosigue Otto] el modelo del pandemónium deja de lado un nivel o un estadio del proceso. Lo que le falta a su modelo no es una proyección en el «espacio fenoménico» del Teatro Cartesiano —¡qué idea más ridícula!—, sino un nivel de articulación adicional en la psicología del hablante. No basta con que las palabras se unan las unas con las otras en una danza nupcial interna para ser proferidas acto seguido. Si tienen que ser *relaciones* de los estados mentales de alguien, deben estar basados, de un modo u otro, en un acto de *aprehensión* interna. Lo que ignora el modelo del pandemónium es el estado de conciencia del sujeto que guía el habla.

Tanto si Otto está en lo cierto como si no, sin duda se está haciendo eco de una visión muy común: así es precisamente como solemos concebir nues-

tra capacidad de contar a las personas nuestros estados conscientes. En una reciente serie de artículos, el filósofo David Rosenthal (1986, 1989, 1990a, b) ha analizado este concepto ordinario de conciencia en relación con nuestros conceptos de *relación* y de *expresión*. Rosenthal descubre algunos rasgos estructurales que nos pueden ser muy útiles. En primer lugar, podemos utilizar su análisis para ver, desde el interior, cuál es la imagen estándar y por qué resulta tan atractiva. En segundo lugar, podemos mostrar cómo desacredita la idea de los zombíes, sin ninguna ayuda exterior. En tercer lugar, podemos volver la imagen estándar contra sí misma, y utilizar las dificultades con que nos encontremos para motivar una nueva visión, una visión que conserva lo bueno de la imagen tradicional pero descarta el marco cartesiano.

Dibujo de Gahan Wilson © 1990 The New Yorker Magazine, Inc. Reproducido con autorización

Noreldo, la maravilla mental, lee la mente de su gato, Ned.

Figura 10.7

¿Qué ocurre cuando hablamos? En el centro de nuestra concepción diaria de este acto hay una perogrullada: en el supuesto de que no mintamos o no estemos siendo sinceros, decimos lo que pensamos. En otras palabras, expresamos una de nuestras creencias o uno de nuestros pensamientos. Su-

ponga, por ejemplo, que ve a su gato parado ansiosamente junto a la nevera y que usted dice «este gato quiere cenar». Esto expresa su creencia de que el gato quiere cenar. Al *expresar* su creencia, usted está *refiriendo* lo que considera un hecho sobre su gato. En este caso usted está refiriendo el deseo del gato de ser alimentado. Es importante señalar que usted no está *refiriendo* su creencia o *expresando* el deseo del gato. El gato expresa su deseo al permanecer parado ansiosamente junto a la nevera, y usted, al percatarse, utiliza este dato como la base —la evidencia— para su relato. Hay muchas maneras de expresar un estado mental (como un deseo), pero sólo hay una de referirlo, es decir, profiriendo un acto de habla (oral, escrito o a través de otro medio).

Una de las maneras más interesantes de *expresar* un estado mental es *refiriendo* otro estado mental. En el ejemplo, usted refiere el deseo del gato y así expresa su propia creencia sobre el deseo del gato. Su conducta evidencia no sólo que el gato tiene un deseo, sino también que usted cree que el gato tiene un deseo. Sin embargo, usted podría haber evidenciado su creencia de otra manera, quizá levantándose de su silla y preparándole la cena al gato. Esto hubiera *expresado* la misma creencia sin *referir* nada. O usted podría haberse quedado sentado y alzar los ojos al cielo, expresando *de forma no intencionada* su exasperación por el deseo del gato justo en el momento en que usted acababa de sentarse cómodamente en su silla. Expresar un estado mental, deliberadamente o no, es hacer algo que evidencia, o pone de manifiesto, ese estado para un observador externo, o un lector de mentes, si usted prefiere. Referir un estado mental, por el contrario, es una actividad más compleja, siempre intencionada y en la que siempre participa el lenguaje.

Aquí tenemos, por tanto, una pista importante sobre los orígenes del modelo del Teatro Cartesiano: nuestra psicología del sentido común de todos los días equipara el referir de los estados mentales de uno con el modelo del referir acontecimientos del mundo exterior. Su relación de que el gato quiere cenar se basa en su observación del gato. Su relación expresa su creencia de que el gato quiere cenar, una creencia sobre el deseo del gato. Llamaremos estados mentales de *segundo orden* a las creencias sobre creencias, a los deseos sobre deseos, a las creencias sobre deseos, a las esperanzas sobre temores, etc. Y si yo (1) *creo* que usted (2) *piensa* que yo (3) *quiero* tomarme un café, esta creencia mía es una creencia de tercer orden. (Sobre la importancia de los estados mentales de orden superior en las teorías de la mente, véase mi libro *The Intentional Stance*, 1987a.) No cabe duda de que existen importantes y notables diferencias marcadas por estas distinciones cuando se aplican de forma no reflexiva, es decir, cuando x cree que y está en un determinado estado mental y $x \neq y$. Hay una diferencia enorme entre el caso en que el gato quiere cenar y usted lo sabe, y el caso en que el gato quiere cenar y usted no lo sabe. ¿Pero qué ocurre con los casos reflexivos, en que $x = y$? La psicología del sentido común trata ambos casos de la misma manera.

Supóngase que refiero que yo quiero comer. En el modelo estándar, debo estar expresando una *creencia de segundo orden* sobre *mi* deseo. Cuando refiero mi deseo, expreso una creencia de segundo orden, mi creencia sobre mi deseo. ¿Qué ocurre cuando refiero esta creencia de segundo orden, diciendo «creo que quiero comer»? Esta relación debe expresar una creencia de *tercer* orden, mi creencia de que en efecto creo que quiero comer. Y así, sucesivamente. Nuestros conceptos usuales sobre lo que significa hablar con sinceridad generan, así, una serie de estados mentales presuntamente distintos: mi deseo es diferente de mi creencia de que tengo un deseo, que a su vez es distinto de mi creencia de que tengo la creencia de tener un deseo, etcétera.

La psicología del sentido común establece también otras distinciones. Como señala Rosenthal (junto a muchos otros), ésta distingue las *creencias*, que son los estados disposicionales subyacentes, de los *pensamientos*, que son estados ocurrentes o episódicos, estados transitorios. Su *creencia de que los perros son animales* ha permanecido continuamente como estado de su mente durante años, pero el llamar yo la atención sobre ella ha engendrado un *pensamiento* en usted, el pensamiento de que los perros son animales, un episodio que sin duda no se habría producido de no mediar mi provocación.

De todo esto se sigue, por tanto, que puede haber pensamientos de primero y segundo orden, y de órdenes superiores: pensamientos sobre pensamientos (sobre pensamientos...). Aquí está, pues, el paso clave: cuando yo expreso una creencia, como por ejemplo la creencia de que quiero comer, no estoy expresando una creencia de orden superior de modo directo; lo que ocurre es que mi creencia subyacente provoca un pensamiento episódico, el *pensamiento* de orden superior de que yo quiero comer, y expreso *este pensamiento* (si decido hacerlo). Según Rosenthal, todas éstas son las implicaciones del modelo de sentido común de *decir lo que uno piensa*.

Dado que un rasgo distintivo de los estados de la conciencia humana es que se pueden referir (salvo en casos de afasia, parálisis, o si uno está atado y amordazado, por ejemplo), se sigue, en el análisis de Rosenthal, que «los estados conscientes deber ir acompañados por los pensamientos de orden superior adecuados, mientras que los estados mentales no conscientes no pueden disfrutar de dicha compañía» (1990b, pág. 16). Evidentemente, el pensamiento de orden superior en cuestión debe ser sobre el estado a que acompaña; debe ser el pensamiento en que uno está (o acaba de estar; el tiempo pasa) dentro del estado de orden inferior. Esto parece que está a punto de generar una regresión infinita de estados conscientes o pensamientos de orden superior, pero Rosenthal argumenta que la psicología del sentido común permite una inversión sorprendente: *El pensamiento de segundo orden no tiene por qué ser consciente para que su objeto de primer orden sea consciente.* Uno puede expresar un pensamiento sin ser consciente de él, de modo que uno puede expresar un pensamiento de *segundo orden* sin ser consciente de *él*, todo lo que necesita es ser consciente de su objeto, el pensamiento de primer orden que uno *refiere*.

En un principio esto puede parecer sorprendente, pero un poco de reflexión nos hará ver que nos hallamos ante un hecho conocido visto desde una nueva perspectiva: uno no presta atención al pensamiento que expresa, sino al objeto (u objetos) *sobre* el que trata el pensamiento. Rosenthal prosigue su argumentación señalando que, aunque algunos pensamientos de segundo orden son conscientes —en virtud de los pensamientos de tercer orden sobre ellos—, éstos son relativamente raros. Son los pensamientos explícitamente introspectivos que referiríamos (aun para nosotros mismos) sólo cuando nos halláramos en estadio de hiperconciencia. Si yo le digo a usted, «siento dolor», estoy refiriendo un estado consciente, mi dolor, y expresando una creencia de segundo orden, mi creencia de que siento dolor. Si, poniéndome filosófico, digo «pienso [o estoy seguro, o creo] que siento dolor», estoy así refiriendo un pensamiento de segundo orden, expresando un pensamiento de tercer orden. Sin embargo, normalmente yo no tendría dicho pensamiento de tercer orden y, por tanto, no sería consciente de dicho pensamiento de segundo orden; lo expresaría diciendo «siento dolor», pero no sería consciente de él.

En un principio, esta idea de los pensamientos inconscientes de orden superior podría parecer disparatada o paradójica, aunque la categoría de los episodios en cuestión no es realmente controvertida, pese a que el término «pensamiento» no suela utilizarse para referirse a ellos. Rosenthal utiliza «pensamiento» como término técnico —siguiendo, de hecho, la práctica de Descartes— que cubre *todos* los estados episódicos con contenido. Así pues, una punzada de dolor o la visión fugaz de unas medias de mujer serían pensamientos para Descartes y para Rosenthal. A diferencia de Descartes, no obstante, Rosenthal insiste en la existencia de pensamientos inconscientes.

Los pensamientos inconscientes serían, por ejemplo, acontecimientos perceptivos inconscientes, o activaciones episódicas de creencias, que ocurren naturalmente —que *deben* ocurrir— en el curso del control normal de la conducta. Suponga que usted derrama una taza de café sobre su mesa de trabajo. En un relámpago, usted salta de la silla, evitando por poco el café que cae por el borde de la mesa. Usted no era consciente del pensamiento de que la superficie de la mesa no iba a absorber el café, ni de que el café, al ser un líquido que obedece la ley de la gravedad, se derramaría por el borde de la mesa y, sin embargo, dichos pensamientos inconscientes deben haberse producido, porque si la taza hubiera estado llena de sal o si la mesa hubiese estado cubierta con un mantel, usted no habría saltado. De todas sus creencias —sobre el café, sobre la democracia, sobre el béisbol, sobre el precio del té en China—, sólo éstas y algunas más se hicieron inmediatamente relevantes para sus circunstancias. Si las citamos en una explicación de por qué usted saltó de la silla, es porque por un instante deben de haber sido accesibles o activadas, o, de un modo u otro, deben de haber sido interceptadas para contribuir a su conducta, aunque, evidentemente, todo eso se produjo de forma inconsciente. Estos episodios inconscientes serían ejemplos de lo que Rosenthal denomina pensamientos inconscientes. (Ya hemos encontra-

do pensamientos inconscientes en ejemplos anteriores: por ejemplo, la percepción inconsciente de las vibraciones en el dedo que le permiten identificar, conscientemente, las texturas tanteadas con una varilla, el recuerdo inconsciente de una mujer con gafas, que llevó a la experiencia errónea de que ésa era la mujer que había pasado.)

Rosenthal señala que al hallar un modo de definir la conciencia en términos de estados mentales inconscientes (los pensamientos de orden superior que acompañan a los otros), ha descubierto un modo de sentar las bases *dentro de la psicología del sentido común* para una teoría de la conciencia que no sea ni circular ni misteriosa (1990b). Argumenta que lo que distingue un estado consciente de un estado inconsciente no es alguna propiedad intrínseca e inexplicable, sino la simple propiedad de poseer un pensamiento de orden que lo acompañe y que trata del estado en cuestión. (Véase Harnad, 1982, donde se utiliza una estrategia similar con variantes bastante interesantes.) Es un buen signo para la psicología del sentido común: ya no está sumida en el misterio; tiene los recursos, bien explotados por Rosenthal, para articular una explicación de su categoría más preciada, la conciencia, en términos de sus categorías subsidiarias y menos problemáticas. Parte del buen negocio, si es que optamos por su análisis, es que se puede utilizar para romper la supuesta distinción que existe entre los seres conscientes y los zombíes.

4. Zombíes, zimbos y la ilusión del usuario

> La mente es una regularidad percibida por una mente. Puede que esto sea circular, pero no es vicioso ni paradójico.
>
> DOUGLAS HOFSTADTER, 1981, pág. 200

Como recordará el lector, los zombíes de los filósofos parecen ejecutar actos de habla, parecer referir sus estados de conciencia, parecen llevar a cabo actos de introspección. Y, sin embargo, no son realmente conscientes, a pesar del hecho de que es imposible distinguir su comportamiento del de una persona consciente. Pueden poseer estados internos con contenido funcional (el tipo de contenido que los funcionalistas pueden asignar a los mecanismos internos de un robot), pero éstos son estados inconscientes. Shakey, según lo imaginamos, es un zombie paradigmático. Cuando «refiere» un estado interno, lo que se refiere no es un estado consciente, ya que Shakey no tiene estados conscientes, sino un estado inconsciente que no tiene otro efecto que el de hacerle entrar en un nuevo estado inconsciente que dirige el proceso de generar y ejecutar lo que podríamos denominar un acto de habla consistente en una fórmula «enlatada». (Hemos dejado que Otto viniera insistiendo en ello desde el principio.)

Shakey no decidió *primero* qué *referir*, después de observar que algo se estaba produciendo en su interior, y decidió *después cómo expresarlo*; Sha-

key simplemente se encontró con cosas que decir. Shakey no tenía acceso al motivo por el cual quería decir que estaba elaborando dibujos lineales alrededor de las fronteras de claro-oscuro de sus imágenes mentales; estaba construido de esta manera. La afirmación principal del capítulo 8 fue, no obstante, que, a pesar de las primeras apariencias, lo dicho vale también para usted. Usted no tiene ningún acceso especial a los motivos por los cuales quiere decir lo que usted encuentra que quiere decir; usted está construido de esta manera. Pero, a diferencia de Shakey, usted está reconstruyéndose a sí mismo constantemente, descubriendo nuevas cosas que quiere decir como resultado de una reflexión sobre lo que usted se ha encontrado queriendo decir, y así sucesivamente.

¿Pero, podría un Shakey más elaborado hacer esto también? Shakey era un zombie bastante burdo, pero ahora podemos imaginar un zombie más realista y complejo, capaz de supervisar sus propias actividades, incluidas sus propias actividades internas, en una indefinida espiral de reflexividad ascendente. Denominaré *zimbo* a esta entidad pensante. Un zimbo es un zombie que, gracias al autocontrol, posee estados informativos internos (pero inconscientes) de orden superior sobre sus otros estados informativos de orden inferior. (No importa para este experimento mental si el zimbo en cuestión es un robot o una entidad humana, o marciana.) Aquellos que creen en la coherencia del concepto de zombie, deben, sin lugar a dudas, aceptar las posibilidad del concepto de zimbo. Un zimbo no es más que un zombie con una conducta compleja, gracias a un sistema de control que permite la autorepresentación recursiva.

Considérese cómo se comportaría un zimbo en el test de Turing, la famosa propuesta de Alan Turing (1950) de un test operacional para detectar ordenadores pensantes. Un ordenador podrá pensar, afirmaba Turing, si éste es capaz de derrotar con regularidad a un oponente humano en el «juego de la imitación»: los dos jugadores permanecen ocultos de la vista de un juez humano, pero pueden comunicarse con él tecleando mensajes en un terminal de ordenador. El participante humano sólo tiene que intentar convencer al juez de que él o ella es efectivamente un humano, mientras que el participante máquina tiene que intentar hacer lo mismo, es decir, convencer al juez de que es un humano. Si el juez no detecta sistemáticamente cuál es el ordenador, entonces puede considerarse que éste posee la capacidad de pensar. Turing propuso su test con la idea de zanjar la cuestión; para él resultaba obvio que el test era tan difícil de superar que cualquier ordenador que fuera capaz de superarlo debería ser considerado por todo el mundo como un pensador *asombroso*. Pensó que había puesto el listón lo bastante alto como para satisfacer a cualquier escéptico, pero se equivocó. Muchos han argumentado que «la superación del test de Turing» no sería una prueba suficiente de inteligencia, y mucho menos de conciencia. (Véase Hofstadter, 1981b, Dennett, 1985a, y French, 1991, para análisis de los puntos fuertes y débiles del test de Turing y sus críticos.)

Ahora bien, las oportunidades de un zimbo en el test de Turing deberían

ser las mismas que las de cualquier persona consciente, ya que los participantes no muestran al juez otra cosa que su conducta; de hecho, muestran únicamente su conducta verbal (tecleando en un terminal). Suponga, pues, que usted es el juez en un test de Turing y que los (presuntos) actos de habla de un zimbo le han convencido de que éste es consciente. Estos presuntos actos de habla no deberían haberle convencido, *ex hypothesi*, ya que no es más que un zimbo, y los zimbos no son conscientes. ¿Debería haberse convencido a sí mismo, sin embargo? Cuando un zimbo profiere un testimonio, expresando un estado inconsciente de segundo orden, no hay nada que le impida reflexionar (inconscientemente) sobre esta situación concreta. De hecho, si tiene que ser convincente, va a tener que ser capaz de responder de manera apropiada a sus propias «afirmaciones» dirigidas a usted (o tener conocimiento de ellas).

Suponga, por ejemplo, que el zimbo es un Shakey perfeccionado y que usted, como juez, acaba de pedirle que resuelva un problema en su ojo de la mente y que le explique acto seguido cómo lo ha hecho. Éste reflexiona sobre la afirmación que acaba de hacerle de que resolvió el problema construyendo un dibujo lineal sobre una imagen mental. Así «sabría» que esto es lo que había querido decir y, si reflexionara un poco más, llegaría a «saber» que no sabía por qué esto era lo que quería decir. Cuanto más le preguntara usted acerca de lo que sabía y lo que no sabía sobre lo que estaba haciendo, más pensativo se haría. Hasta aquí, paree que hemos conseguido imaginar un ser inconsciente que sin embargo posee la capacidad de tener pensamientos de orden superior. Pero, de acuerdo con Rosenthal, cuando un estado mental va acompañado de un pensamiento de orden superior consciente *o* inconsciente a fin de que dicho estado mental pueda existir, ¡ello garantiza *ipso facto* que el estado mental sea un estado consciente! ¿Nuesto experimento mental desacredita el análisis de Rosenthal o desacredita la definición de zimbo?

Es fácil comprobar que como mínimo el zimbo creería (inconscientemente) que se halla en diversos estados mentales, precisamente los estados mentales que su posición le permitiría referir si nosotros le hiciéramos preguntas. *¡Pensaría* que es consciente, aunque no lo fuese! Cualquier entidad que pudiera superar el test de Turing operaría bajo la sensación (¿errónea?) de que es consciente. En otras palabras, sería víctima de una ilusión (véase Harnad, 1982). ¿Qué tipo de ilusión? Una ilusión de usuario, por supuesto. ¡Sería «víctima» de una ilusión benigna producida por su propia máquina virtual!

¿No es esto un juego de espejos, una especie ilegítima de juego de manos de filósofos? ¿Cómo es posible que haya una ilusión del usuario sin que haya un Teatro Cartesiano en el que se perpetre la ilusión? Parece como si me hallara ante el inminente peligro de verme derrotado por mis propias metáforas. El problema es que la ilusión del usuario de una máquina virtual *se logra* gracias a la presentación de material en una especie de teatro, y que hay una audiencia externa e independiente, el usuario, en cuyo beneficio se

lleva a cabo el espectáculo. En este momento estoy utilizando un ordenador, tecleando estas palabras en un «archivo» con la discreta ayuda de un programa de procesamiento de textos. Cuando interactúo con el ordenador, tengo un acceso limitado a los acontecimientos que se producen en su interior. Gracias a los esquemas de presentación ideados por los programadores, me veo invitado a participar de una elaborada metáfora audiovisual, una obra interactiva representada en el escenario del teclado, el ratón y la pantalla. Yo, el usuario, estoy sujeto a toda una serie de ilusiones benignas: parezco ser capaz de mover el cursor (un siervo visible y poderoso) hacia el lugar dentro del ordenador donde guardo mi archivo y, cuando veo que el cursor ha llegado hasta «allí», presionando una tecla recupero el archivo que, como un largo pergamino, se desenrosca frente a una ventana (la pantalla) bajo mi control. Tecleando diferentes mandos y presionando diversos botones, puedo hacer que ocurra todo tipo de cosas en el interior de mi ordenador sin que yo tenga que conocer los detalles; mantengo el control con sólo confiar en mi comprensión de las detalladas metáforas audiovisuales provistas por la ilusión del usuario.

Para la mayoría de los usuarios de ordenadores, es sólo en estos términos que se llega a apreciar lo que está ocurriendo en el interior de la máquina. Éste es uno de los hechos que hace de la máquina virtual una analogía tan tentadora para la conciencia, porque siempre ha parecido que nuestro acceso a lo que está ocurriendo en el interior de nuestros cerebros es limitado; *nosotros* no tenemos por qué saber de qué manera la maquinaria entre bastidores del cerebro lleva a cabo su magia; *nosotros* tenemos conocimiento de sus operaciones únicamente en la medida de que éstas se nos presentan ataviadas con las metáforas interactivas de la fenomenología. Pero si, cuando echamos mano de esta tentadora analogía, mantenemos esa «obvia» separación entre presentación por un lado y apreciación del espectáculo por parte del usuario por el otro, parece que volvemos irremediablemente al Teatro Cartesiano. ¿Cómo es posible que haya una ilusión del usuario sin esta separación?

No podría existir; el usuario que fija la perspectiva desde la cual la máquina virtual se hace «visible» tiene que ser un observador externo, un *Vorsetzer*. En un principio, uno podría pensar que este concepto de observador externo debe coincidir con el concepto de observador consciente, pero ya hemos visto que no tiene por qué ser necesariamente así. El Vorsetzer que se sentaba frente al sistema de CAD en el sistema CADBLIND Mark I original no era consciente, pero tenía un acceso tan limitado al funcionamiento interno del sistema de CAD como el que tendría cualquier usuario consciente. Y, una vez que hemos descartado el material innecesario formado por la pantalla y la cámara, la presentación y la apreciación del usuario se evaporan, sustituidas, como ya ha ocurrido muchas veces en nuestra explicación, por una serie de modestas transacciones. El «observador externo» puede irse incorporando gradualmente en el sistema, dejando algunas huellas fósiles: retazos de «interficie» cuyos variados formatos siguen restringiendo

el tipo de preguntas que se pueden responder y, así, restringen también los contenidos que se pueden expresar.[8] No tiene por qué haber un único lugar donde se produzca la presentación.[9] Y, según nos sugiere el análisis de Rosenthal, incluso nuestra noción ordinaria de conciencia, anclada en esas intuiciones del sentido común de la psicología del sentido común, puede tolerar la inconciencia de los estados de orden superior cuya presencia en el sistema da cuenta de la conciencia de algunos de sus estados.

¿Es, pues, el proceso de la reflexión inconsciente el camino a través del cual un zombie se puede convertir en un zimbo y hacerse, así, consciente? Si lo es, entonces los zombíes deber de ser conscientes después de todo. Todos los zombíes son capaces de proferir «actos de habla» convincentes (recuérdese que son indistinguibles de nuestros mejores amigos), y esta capacidad sería mágica si las estructuras de control o los procesos causalmente responsables de dicha capacidad en el cerebro del zombie (o del ordenador o de cualquier otra cosa) no fueran acompañadas de la reflexión suficiente sobre los actos y su contenido (aparente o funcional). Un zombie podría empezar su carrera en un estado no comunicativo y no reflexivo y ser, así, un verdadero zombie, un ser inconsciente, pero tan pronto como empezara a «comunicarse» con otros y consigo mismo, empezaría a estar equipado con el mismo tipo de estados que, según el análisis de Rosenthal, bastan para la conciencia.

Si, por otra parte, rechazamos el análisis de la conciencia de Rosenthal en términos de pensamientos de orden superior, entonces los zombíes podrán seguir viviendo hasta el día del próximo experimento mental. Ofrezco esta parábola de la lengua de los zimbos con cierto descaro porque creo que *tanto* el concepto de zombie *como* las categorías de pensamiento de orden superior de la psicología del sentido común no podrán sobrevivir más que como reliquias de una vieja ideología. Rosenthal, sin embargo, nos ha hecho un gran favor al exponer la lógica de estos conceptos tradicionales, y gra-

8. En el capítulo 7 (pág. 236), me preguntaba, «¿hacer llegar los fragmentos adecuados de información hasta la superficie de *qué*?» y prometí dar una respuesta más adelante. La superficie (metafórica) viene determinada por el formato de las interacciones entre partes.

9. Es interesante comparar los diferentes vestigios de esta idea del usuario en el cerebro en el trabajo de los más variados pensadores. Aquí están las palabras de Minsky (1985): «Exagerando un poco, lo que denominamos [conciencia] consiste poco más que en listados que se encienden, periódicamente, en pantallas mentales utilizadas por otros sistemas [pág. 57; pág. 59 de la tr. cast. (N. del T.)]... Dividamos el cerebro en dos partes, A y B. Conectemos las entradas y salidas del cerebro A al mundo real, para que pueda percibir lo que ocurre allí. Pero no vinculemos el cerebro B con el mundo exterior; ¡efectuemos, en cambio, conexiones para que el cerebro A sea el mundo del cerebro B!» [pág. 59; pág. 61 de la tr. cast. (N. del T.)]. Con muy buen criterio, Minsky evita postular líneas divisorias anatómicas, aunque otros estudiosos se han apresurado a proponer algunas. Cuando Kosslyn especuló por primera vez sobre la conciencia en tanto que máquina virtual, se inclinó por situar al usuario en los lóbulos frontales (véase también Kosslyn, 1980, pág. 21), y, más recientemente, Edelman, siguiendo su propia línea de argumentación, ha llegado a la misma conclusión, planteada en términos de su «memoria de yo/no yo dominada por valores» que él sitúa en los lóbulos frontales y a la cual asigna la tarea de interpretar las producciones del resto del cerebro (Edelman, 1989, págs. 102 y sigs.).

cias a la clara visión que ahora tenemos de ellos, podemos percibir también con mejor claridad cómo serían sus propios sustitutos.

5. Problemas con la psicología del sentido común

Rosenthal considera que la psicología del sentido común postula una jerarquía expandible de pensamientos de orden superior, concebidos como episodios destacados, independientes y llenos de contenido que se producen en la mente en tiempo real. ¿Cómo se sostiene esta visión cuando intentamos confirmarla? ¿Existen estos estados y acontecimientos distintos en el cerebro? Si somos generosos en el momento de definir lo que cuenta, la respuesta tiene que ser sí. No cabe duda de que existen ciertas conocidas diferencias psicológicas que pueden —y suelen— ser descritas en estos términos.

> «De repente, Dorothy se dio cuenta de que quería marcharse, y de que hacía mucho rato que quería hacerlo.»

Parece que aquí Dorothy adquirió la creencia de segundo orden —tras tener un pensamiento de segundo orden— sobre su deseo algún tiempo después de que el deseo hubiera surtido efecto. Existen numerosos casos cotidianos de estos tipos: «Y entonces se dio cuenta de que estaba mirando precisamente hacia donde estaba el gemelo que le faltaba». «La quiere, pero todavía no se ha dado cuenta.» Resulta difícil negar que estos enunciados de la vida diaria aluden a transiciones genuinas de un «estado mental» a otro. Y como señala Rosenthal, intuitivamente esa transición consiste en *hacerse consciente* del estado de primer orden. Cuando Freud, al basarse en estos casos cotidianos, postuló ese vasto reino de estados mentales inconscientes, se trataba precisamente de esos estados en los que sus sujetos no creen estar. Estas personas están en estados mentales en los que todavía no se les ha ocurrido —a través de pensamientos de orden superior— que están.

Esta manera de describir dichas diferencias resulta familiar, aunque decidir si es adecuada o no ya es otra cuestión. Todas éstas son transiciones hacia un estado mejor informado (por hablar de la manera más neutral posible), y estar mejor informado es condición necesaria para *referir* (en contraposición a *expresar*) el «estado mental» anterior. Ahora bien, sería pueril describir este proceso de la siguiente manera: *a fin de* referir un estado o un acontecimiento mental, es preciso tener un pensamiento de orden superior que expresar. Ello comporta la imagen de primero *observar* (con algún órgano sensorial interno) el estado o acontecimiento mental, produciendo *así* un estado de *creencia*, cuyo inicio está marcado por un *pensamiento*, que acto seguido se expresa. Esta cadena causal, como vimos, simula la cadena causal de referir acontecimientos externos ordinarios: usted primero observa el acontecimiento con la ayuda de sus órganos sensoriales, lo cual produce

una creencia en usted, y después un pensamiento, que usted expresa en su relación.

Este hipotético pensamiento de orden superior es, creo, el «nivel de articulación adicional» que Otto creía poder discernir en su propia psicología; es el pensamiento que *expresan* las palabras de Otto cuando *refiere* su propia experiencia consciente. Pero según el modelo de generación del habla que esbozamos en el capítulo 8, el modelo de Otto tiene el orden causal invertido. No entramos *primero* en un estado de autoobservación de orden superior, creando un pensamiento de orden superior, de modo que uno puede entonces referir el pensamiento de orden inferior expresando el pensamiento de orden superior. Por el contrario, el estado de segundo orden (el estado mejor informado) es *creado* por el proceso mismo de construir la relación. No aprehendemos *primero* nuestra experiencia en el Teatro Cartesiano y *después*, a partir de ese conocimiento adquirido, tenemos la habilidad de construir testimonios y expresarlos; nuestro *ser capaces de decir* lo que se siente *es la base de* nuestras «creencias de orden superior».[10]

En un principio, un proceso del diseño de actos de habla en pandemónium parece equivocado porque parece dejar de lado al observador/decididor central cuyo pensamiento será finalmente expresado. Pero ésta es una ventaja, no un inconveniente, de este modelo. La emergencia de la expresión es precisamente lo que crea o fija el contenido del pensamiento de orden superior expresado. No es necesario que haya ningún «pensamiento» episódico *adicional*. El estado de primer orden depende materialmente —causalmente— de la expresión del acto de habla, pero no necesariamente de la pública expresión de un acto de habla explícito. En el capítulo 7 vimos de qué modo las necesidades que tenía un organismo de poseer una comunicación interna de la información cada vez mejor podrían haber conducido a la creación de hábitos de automanipulación que pudieron ocupar el lugar del proceso evolutivo más laborioso de crear un ojo interno, un órgano interno real que pudiera controlar al cerebro. Conjeturamos que la única manera que tiene el cerebro de entrar en algo parecido a un estado de creencia de orden superior pasa por ocuparse en un proceso muy parecido a referirse estados de primer orden a sí mismo.

10. Esto es, cuando menos, un pariente cercano del tema central de los trabajos últimos de Wittgenstein, aunque el propio Wittgenstein renunció a desarrollar una explicación positiva o modelo de la relación entre aquello que decimos y aquello de lo que estamos hablando cuando referimos (aparentemente) nuestros estados mentales. La filósofa Elizabeth Anscombe, en su clásico y frustrantemente oscuro libro, *Intention* (1957), intentó llenar este vacío dejado por Wittgenstein, argumentando que es erróneo asegurar que *sabemos* cuáles son nuestras intenciones; por el contrario, sólo *podemos decir* cuáles son nuestras intenciones. También intentó caracterizar la categoría de las cosas que podemos *conocer sin observación*. En mi libro *Content and Consciousness* (1969), capítulos 8 y 9, se desarrolla una discusión bastante imperfecta de estas afirmaciones de Anscombe. Siempre he creído, sin embargo, que ahí había algo de razón, oculta en aquellos pasajes. Mi segundo intento se halla en «Toward a cognitive theory of consciousness» (1978) reimpreso en *Brainstorms*, especialmente las secciones IV y V (págs. 164-171). Este apartado es mi actual intento de aclarar estas ideas y constituye, además, una orientación radicalmente diferente de las versiones anteriores.

Debemos acabar con el hábito de postular observadores cada vez más centrales. A modo de transición, podemos reconsiderar el proceso no como un modelo de conocimiento por observación sino como un modelo de *rumores*. Creo que *p* porque una fuente fiable me ha dicho que *p*. ¿Qué fuente? Yo mismo o, en todo caso, uno de mis «agentes». Ésta no es una manera tan extraña de pensar; después de todo, hablamos del *testimonio* de nuestros sentidos, una metáfora que sugiere que nuestros sentidos no aportan pruebas materiales ante el «tribunal» para *mostrárnoslas*, sino que nos *hablan* de cosas. Apoyándonos en esta metáfora (hasta que nos hayamos acostumbrado a las complejidades de una alternativa mejor), podemos proponer el siguiente eslogan:

> «Si no pudiera hablar conmigo mismo, no tendría ninguna manera de saber lo que estoy pensando».

Ésta no es todavía la mejor manera de pensar en ello, por muchos motivos. Primero, existe una diferencia —que he venido ignorando hasta el momento— entre una entidad «hablando consigo misma» y una serie de subsistemas «hablando los unos con los otros». En el capítulo 13, donde tratamos el problema del yo, desarrollamos una transición adecuada entre ambas ideas. Segundo, como hemos visto, el hincapié en la expresión lingüística es una exageración; hay otras estrategias de automanipulación y autoexpresión que no son verbales.

Podría parecer que estoy proponiendo un mal negocio: abandonar la relativa claridad y transparencia del modelo estándar de la psicología del sentido común, con su jerarquía de observaciones interiores, en favor de una alternativa incompleta que apenas podemos concebir. Pero la claridad del modelo tradicional no es más que una ilusión, por motivos que ya apuntábamos en el capítulo 5, cuando examinamos el extraño caso de la *apariencia real*. Ahora podemos hacer un diagnóstico más preciso de los problemas. Otto es el portavoz de la psicología del sentido común y, si le dejamos continuar, se hará pronto un lío. La visión de Otto, que tenazmente extiende las categorías de la psicología del sentido común «hacia adentro», genera una explosión de «estados representacionales» distintos, cuyas interrelaciones generan problemas artificiales. Otto continúa:

> Mi testimonio público de un estado consciente, en caso de que decida hacerlo, podría contener un error. Podría tener un lapsus, o desconocer el significado exacto de una palabra y, así, malinformale a usted de forma totalmente involuntaria. Cualquiera de estos errores de expresión que me pasara por alto podría crearle una creencia falsa sobre los hechos, sobre la manera en que yo lo siento *realmente*. Y el mero hecho de que yo no consiguiera detectar un error, no significaría que no hubiera un error. *Por una parte*, está la verdad sobre lo que yo siento, y, *por otra parte*, está lo que yo finalmente digo sobre lo que yo siento (si decido hacerlo). Aunque mis relaciones tienden a ser bastante fiables, siempre se puede colar un error.

Ésta es una de aquellas situaciones en que con dos manos no basta. Pues, como Rosenthal nos ha demostrado, además de «lo que yo siento» y «lo que finalmente digo» parece que tiene que haber un tercer hecho intermedio: mi creencia de lo que yo siento.[11] Ya que cuando digo sinceramente lo que digo, queriendo decir lo que quiero decir, expreso una de mis creencias: mi creencia sobre lo que yo siento. Evidentemente, hay todavía un cuarto hecho intermedio: mi pensamiento episódico de que así es como yo lo siento.

¿Podría ser errónea mi creencia sobre lo que yo siento? O ¿no podría ser que yo *sólo piense* que así es como lo siento? O, en otras palabras, ¿no podría sólo *parecerme* que ésta es mi actual experiencia? Otto quería una separación, pero ahora se nos amenaza con más: entre la experiencia subjetiva y la creencia sobre ella, entre la creencia y el pensamiento episódico engendrado por el camino hacia la expresión verbal, y entre ese pensamiento y su expresión última. Y como las escobas del aprendiz de brujo que se multiplican, todavía quedan más en perspectiva una vez hemos aceptado éstas. Supongamos que tengo mi experiencia subjetiva (una), la que sienta las bases para la creencia de que la estoy teniendo (dos), que a su vez engendra el pensamiento asociado (tres) que después suscita en mí una intención comunicativa por expresarlo (cuatro), que, finalmente, da lugar a la expresión real (cinco). ¿No hay acaso espacio para que se cuelen errores en la transición entre cada una de ellas? ¿No podría ser que yo crea una proposición pero, debido a una transición errónea entre estados, llegue a *pensar* una proposición distinta? (Si uno puede «malhablar», ¿no podrá acaso también «malpensar»?) ¿No sería posible construir la intención de expresar una proposición totalmente diferente de la que se está pensando? Y ¿no podría un recuerdo defectuoso en el subsistema de intenciones comunicativas llevar a disponer un mensaje preverbal para ser expresado y acabar con un mensaje preverbal diferente que funcione como muestra para la corrección de errores? Entre dos cosas distintas siempre queda espacio lógico para los errores y, a medida que multiplicamos los estados con contenidos definidos, descubrimos —o creamos— múltiples fuentes de error.

Se puede tener la tentación de cortar este nudo gordiano, declarando que mi *pensamiento (o mi creencia) sobre lo que yo siento es exactamente la misma cosa que lo que es realmente mi experiencia.* En otras palabras, se tiene la tentación de insistir en el hecho de que no hay espacio lógico para que se cuele ningún error entre ellos, ya que son la misma cosa. Esta afirmación tiene algunas propiedades interesantes. Detiene la amenazadora explosión en el primer paso —que, normalmente, suele ser el mejor lugar para parar una explosión o una regresión— y tiene cierto atractivo desde el punto de vista intuitivo, que podemos expresar con una pregunta retórica: ¿Qué sentido le podemos dar a la afirmación de que algo me pareció que me parecía (que me parecía...) ser un caballo?

11. En *Brainstorms*, exploté este rasgo de la psicología del sentido común en mi examen de «las multiplicidades-β de creencias fenomenológicas» (1978a, págs. 177 y sigs.)

Pero aquí tenemos que pisar con cuidado, andando de puntillas entre los esqueletos de teorías filosóficas difuntas (incluidas algunas de las mías, véase Dennett, 1969, 1978c, 1979a). Podría parecer que podemos quedarnos con las viejas y pasadas de moda categorías de la psicología del sentido común, es decir, las creencias, los pensamientos, las creencias sobre creencias, los pensamientos sobre experiencias, etc., y evitar las perplejidades del autoconocimiento con sólo aunar los casos reflexivos de orden superior e inferior: afirmando que cuando creo que creo que *p*, por ejemplo, se sigue lógicamente que creo que *p*, y, en la misma línea, cuando pienso que siento dolor, se sigue lógicamente que siento dolor, y, así, sucesivamente.

Si así fuera, cuando yo expresase una creencia de segundo orden refiriendo una creencia de primer orden, por ejemplo, en realidad se trataría solamente de un estado, una sola cosa, y el hecho de que al referir una cosa estuviese expresando «otra» se debería a una mera distinción verbal, como en el caso de que Jones decidió casarse con su novia y acabó casándose con su mujer.

Esta unificación, sin embargo, no nos permite cumplir nuestros objetivos. A fin de ver por qué esto es así, considérese una vez más el papel de la memoria, según la concibe la psicología del sentido común. Incluso si resulta intuitivamente plausible que usted no esté equivocado sobre lo que *siente* en *este momento*, no resulta intuitivamente plausible que usted no esté equivocado sobre lo que *sentía entonces*. Si la experiencia que usted está refiriendo es una experiencia pasada, su memoria —aquello sobre lo que usted se basa para elaborar su relación— podría estar contaminada por algún error. Quizá su experiencia se produjera de una manera determinada, pero usted ahora la recuerda como si se hubiera producido de otra. Evidentemente, *ahora*, a usted podría parecerle que *entonces* le pareció que fue un caballo, aunque, *de hecho, entonces*, a usted le pareció que fue una vaca. La posibilidad lógica de recordar mal siempre queda abierta, por muy breve que sea el intervalo que separa el momento de la experiencia y el momento del recuerdo; esto es precisamente lo que permitió el desarrollo de las teorías orwellianas. Pero, como vimos en el capítulo 5, el error que se cuela en las creencias posteriores gracias a la falsificación orwelliana de recuerdos es indistinguible —tanto desde fuera como desde dentro— del error que se cuela en la experiencia original gracias a la construcción de ilusiones estalinianas. Así pues, incluso si pudiéramos mantener que usted tiene un acceso «directo» e «inmediato» a sus *juicios* actuales (su pensamiento de segundo orden sobre lo que usted siente), usted no será capaz de eliminar la posibilidad de que sea un *mal* juicio sobre lo que sintió hace sólo un momento.

Si individuamos los estados (creencias, estados de conciencia, estados de intención comunicativa, etc.) por su contenido —que es la manera tradicional de individuar en la psicología del sentido común—, acabamos por vernos obligados a postular diferencias que son sistemáticamente imposibles de descubrir por cualquier medio, desde dentro o desde fuera y, durante el proceso perdemos la intimidad subjetiva o la incorregibilidad que se supo-

ne que es el sello de la conciencia. Ya pudimos ver ejemplos de ello en el capítulo 5 en nuestro examen de los modelos orwellianos y estalinianos de los fenómenos temporales. Y la solución no consiste en aferrarse a una u otra doctrina que nos proporcione la psicología del sentido común, sino en abandonar esta característica de la psicología del sentido común.

Sustituimos esta división en *estados* discretos y llenos de contenido —creencias, metacreencias, etc.— por un proceso que sirve, a la larga, para asegurar un buen ajuste entre los acontecimientos internos portadores de información de una entidad y la capacidad de esa entidad para expresar verbalmente (parte de la) información de esos acontecimientos. Esto es precisamente lo que se suponía (Otto, por ejemplo) que tenían que asegurar los estados de orden superior, pero no consiguieron llegar hasta el corazón de la naturaleza. De hecho, postulaban corazones que eran sistemáticamente indiscernibles en la naturaleza.

Estos artefactos de la psicología del sentido común sobreviven, sin embargo, como habitantes en los mundos heterofenomenológicos de los sujetos cuyas visiones del mundo están conformadas por este esquema conceptual. Por expresarlo de forma tautológica, ya que a las personas realmente les parece que tienen tanto estas creencias sobre sus experiencias como (además) las experiencias mismas, dichas experiencias y creencias sobre las experiencias son ambas parte de lo que ellas sienten. Así que lo que tenemos que explicar es *este* hecho, no el hecho de que nuestras mentes están organizadas en jerarquías de «estados» representacionales de orden superior de creencias, metacreencias, etc., sino que *tiende a parecernos* que nuestras mentes están ordenadas así.

He aventurado dos motivos por los cuales tendemos a considerar que esta idea es tan atractiva. Primero, persistimos en el hábito de postular un proceso independiente de observación (ahora de observación *interna*) entre las circunstancias sobre las cuales podemos emitir construcciones y las construcciones que finalmente emitimos, pasando por alto el hecho de que en algún momento debemos detener esa regresión de observadores internos con un proceso que unifica los contenidos con sus expresiones verbales sin que medie ningún apreciador de contenidos. Y segundo, las comunicaciones internas creadas de este modo sí que tienen el efecto de organizar nuestras mentes en sistemas de reflexión y de autocontrol infinitamente poderosos. A menudo se ha afirmado, y con razón, que dichos poderes de reflexión están en el centro de la conciencia. Podemos utilizar el simplísimo modelo de la psicología del sentido común como una especie de muleta para la imaginación cuando intentamos comprender los sistemas de autocontrol, pero cuando lo utilizamos, nos arriesgamos a caer en el materialismo cartesiano. Deberíamos ir aprendiendo a seguir adelante sin él, y en el próximo capítulo seguiremos avanzando en esta dirección.

Capítulo 11

DESMANTELANDO EL PROGRAMA DE PROTECCIÓN DE TESTIGOS

1. Resumen.

En la primera parte, examinamos los problemas y establecimos algunos presupuestos y principios metodológicos. En la segunda parte, esbozamos un nuevo modelo de la conciencia, el modelo de las Versiones Múltiples, e iniciamos la tarea de demostrar por qué es preferible al modelo tradicional, el Teatro Cartesiano. Mientras que la idea del Teatro Cartesiano, una vez hecha explícita, exhibe sus errores con gran claridad —no existe ningún materialista cartesiano confeso—, los supuestos básicos y los hábitos de pensamiento que ésta ha fomentado siguen motivando objeciones y distorsionando la «intuición». Ahora, en la tercera parte, nos disponemos a investigar las implicaciones de nuestro modelo alternativo, respondiendo a toda una serie de apremiantes objeciones. Algunas de estas objeciones traicionan una persistente —aunque negada— alianza con el viejo y querido Teatro Cartesiano.

«Pero, ¿dónde se produce la comprensión?» Ésta es la pregunta que ha permanecido oculta en el centro de la controversia desde el siglo XVII. Descartes se encontró con un muro de escepticismo cuando insistió (con razón) en el hecho de que los mecanismos cerebrales podían explicar una buena parte del proceso de comprensión. Por ejemplo, Antoine Arnauld, en sus objeciones a las *Meditaciones*, señaló que «a primera vista parece increíble que pueda ocurrir, sin la asistencia de un alma, que la luz reflejada por el cuerpo de un lobo en los ojos de una oveja mueva las diminutas fibras de los nervios ópticos, y que al alcanzar el cerebro, este movimiento esparza los espíritus animales por los nervios de modo que precipite la huida de la oveja» (1641, pág. 144). Descartes respondió que esto no es más increíble que la capacidad que tienen los humanos de soltar los brazos para protegerse al caer, una reacción que también se produce de forma completamente mecánica, sin la ayuda de un «alma». Esta idea de una interpretación «mecánica» en el cerebro es la base de *toda* teoría materialista de la mente, pero constituye un desafío para una intuición con mucho arraigo: nuestro sentido de que para que se produzca un entendimiento *real*, debe haber *alguien ahí* para validar los procedimientos, para *presenciar* los acontecimientos cuyo curso constituye el entendimiento. (El filósofo John Searle explota esta intuición

en su famoso experimento mental de la Habitación/Sala/Cámara China, que consideraremos en el capítulo 14.)

Descartes era el mecanicista *par excellence* cuando se trataba de cualquier otro fenómeno de la naturaleza, pero cuando se trataba de la mente humana, se arredraba. Además de una interpretación mecánica, afirmaba Descartes, el cerebro también proporciona material para un ruedo central —lo que yo he venido llamando el Teatro Cartesiano— donde, en los seres humanos, el alma puede ejercer de testigo y emitir así sus propios juicios. Los testigos necesitan de materias primas sobre las que basar sus juicios. Estas materias primas, se denominen «datos sensoriales», «sensaciones», «sentimientos brutos» o «propiedades fenoménicas de la experiencia», son los puntales sin los que la idea de un testigo carece por completo de sentido. Estos puntales, firmemente asentados gracias a diversas ilusiones, rodean la idea de un testigo central con una barrera de intuiciones casi impenetrable. Nuestra tarea en este capítulo consiste en derribar esta barrera.

2. Pacientes con visión ciega: ¿zombíes parciales?

De todos los accidentes terribles que afectan a las personas, sólo una pequeña fracción se ve redimida, en parte, por el hecho de que revelan algo de los misterios de la naturaleza a inquisitivos científicos. Esto es particularmente cierto en el caso de las lesiones cerebrales producidas por algún traumatismo (heridas de bala, accidentes de tráfico, etc.), un tumor o un ataque.[1] Las formas de discapacidad y las capacidades mermadas que resultan evidencian, a veces de forma sorprendente, de qué modo el cerebro produce la mente. Una de las más sorprendentes, como su paradójico nombre indica, es la visión ciega. Parace haber sido creada a la medida de los experimentos mentales de los filósofos: una afección que convierte a una persona normal y consciente en un zombie parcial, un autómata inconsciente por lo que a ciertos estímulos respecta, pero una persona consciente normal para el resto. Así que no debe sorprendernos que los filósofos hayan elevado la visión ciega al rango de mito, un ejemplo sobre el cual construir argumentos. Como veremos, sin embargo, la visión ciega no avala el concepto de zombie; al contrario, lo socava.

En la visión humana normal, las señales de entrada viajan desde los ojos por el nervio óptico a través de diversas etapas hasta el *córtex occipital* o córtex visual, la parte del cerebro situada en la parte posterior del cráneo, justo encima del cerebelo. La información sobre el campo visual izquierdo (las mitades izquierdas de los campos de cada ojo) se reparte por el córtex

1. Tim Shallice, en *From Neuropsychology to Mental Structure* (1988), presenta una discusión puesta al día y bien argumentada del razonamiento implicado en el análisis de estos experimentos de la naturaleza. Muchos libros recientes incluyen presentaciones accesibles de algunos de estos casos tan fascinantes: Howard Gardner, *The Shattered Mind* (1975), y Oliver Sacks, *The Man Who Mistook His Wife for His Hat* (1985).

visual derecho, y la del campo visual derecho se reparte por el izquierdo. Ocasionalmente, un accidente vascular (por ejemplo, la ruptura de un vaso sanguíneo) destruye una porción del córtex occipital, creando un punto ciego o *escotoma*, un hueco relativamente grande en el mundo que se experimenta visualmente, en el lado opuesto donde se produjo la lesión.

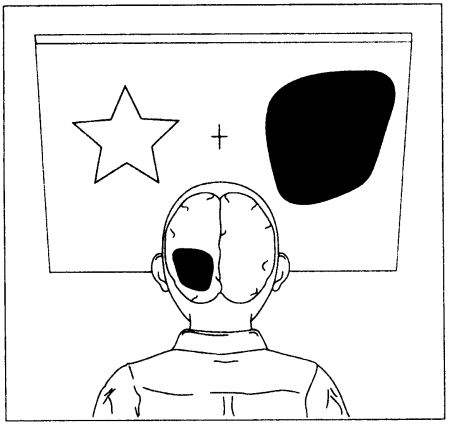


Figura 11.1

En el caso extremo en que tanto el córtex izquierdo como el derecho han sido destruidos, la persona queda ciega por completo. Más comúnmente, todo el córtex visual en un lado del cerebro es destruido por un accidente vascular, provocando la pérdida de la mitad opuesta del campo visual; la pérdida del córtex visual izquierdo produciría una *hemianopsia derecha*, ceguera completa del hemicampo derecho.

¿Qué se siente cuando se tiene un escotoma? Podría parecer que es algo familiar para todos nosotros, ya que todos tenemos puntos ciegos en nuestros campos visuales, en correspondencia con las zonas de nuestra retina en que no hay ni conos ni bastoncillos, porque ése es el punto del globo ocular de donde parte el nervio óptico. Un punto ciego normal, o disco óptico, no es pequeño: oscurece un área circular con un diámetro de aproximadamente unos 6 grados de ángulo visual. Cierre un ojo y mire a la cruz de la figura 11.2, manteniendo la página a unos veinticinco centímetros de sus ojos. Uno de los discos de «punto ciego» debería desaparecer. Cierre el otro ojo, y desaparecerá el otro disco. (Puede que necesite ajustar un poco la distancia de la página para que el efecto deseado se produzca. No deje nunca de mirar en dirección a la cruz.) ¿Por qué no percibe normalmente este hueco en su campo visual? En parte porque usted tiene dos ojos, y cada ojo suple las deficiencias del otro; sus puntos ciegos no se solapan. Pero incluso cuando

usted tiene un ojo cerrado, en la mayoría de los casos usted no percibirá su punto ciego. ¿Por qué? Dado que su cerebro nunca tiene que tratar con información de entrada procedente de esta área de la retina, éste no dedica recursos para operar con esa información. No hay ningún homúnculo responsable de recibir información procedente de esa área, así que como no llega información, tampoco hay nadie para quejarse de ello. Una ausencia de información no es lo mismo que información sobre una ausencia. Para que usted pueda percibir un vacío, algo en su cerebro deberá responder a un contraste: *sea* entre la frontera entre el interior y el exterior —y su cerebro carece de los mecanismos para hacer esto en este punto— *o* entre el antes y el después: ahora ve el disco, ahora no lo ve. (Así es como la desaparición del disco negro de la figura 11.2 le advierte de su punto ciego.)

Figura 11.2

Al igual que nuestros puntos ciegos normales, los escotomas poseen localizaciones definidas y algunos tienen incluso unas fronteras claras que puedes ser fácilmente delimitadas por un investigador, utilizando un estímulo como por ejemplo un punto de luz que se va moviendo por el campo visual del sujeto. Se pide al sujeto que avise cuando deja de experimentar el punto de luz, una variante del experimento que usted acaba de llevar a cabo para descubrir su propio punto ciego. Los testimonios del sujeto pueden entonces correlacionarse con mapas de la parte dañada del córtex, producidos con escáners de CT (computer-aided tomography; tomografía asistida por ordenador) y MRI (magnetic resonance imaging; imagen por resonancia magnética) del cerebro. Un escotoma difiere en un aspecto importante de un punto ciego normal: el sujeto suele percatarse de su presencia. Ello no se debe solamente a que sea mucho mayor que un punto ciego normal. Toda vez que su causa es la pérdida de células en el córtex visual que con antelación «informaban a» otras células del córtex que «se interesaban» por la información proveniente de ciertas regiones de las retinas, su ausencia es percibida. Las expectativas del cerebro se ven alteradas; algo que debía estar ahí ahora ya no está, parte del apetito epistémico no se ve satisfecha. Así pues, el sujeto suele ser consciente del escotoma, pero como una pérdida, no como un área positiva en negro, como le ocurriría a usted si alguien pegara un círculo de papel negro sobre el parabrisas de su coche.

Dado que las vías normales de la información visual en el cerebro se han visto alteradas o cortadas, sería de esperar que las personas que padecen de escotoma fuesen totalmente incapaces de captar ninguna información visual sobre lo que ocurre un su campo visual ciego. Después de todo, son ciegos. Y eso es precisamente lo que ellos mismos dicen: no experimentan nada

visual dentro de los límites de sus escotomas: ni destellos, ni siluetas, ni colores, ni puntos luminosos, ni chispazos. Nada. Eso es lo que significa ser ciego. Sin embargo, algunos pacientes con escotoma muestran un insospechado talento: a pesar de su total carencia de experiencias visuales conscientes en la zona ciega, a veces pueden «adivinar» con sorprendente precisión si hubo o no un destello de luz dentro de la zona, e incluso si se les ha mostrado un cuadrado o un círculo. Éste es el fenómeno conocido con el nombre de visión ciega (Weiskrantz, 1986, 1988, 1990). Cómo debe explicarse este fenómeno sigue siendo aún una cuestión controvertida, aunque ningún investigador piensa que se esté produciendo algún fenómeno «paranormal». Hay por lo menos diez vías diferentes entre la retina y el resto del cerebro, de modo que incluso cuando el córtex occipital ha sido destruido, sigue habiendo bastantes canales de comunicación a través de los cuales puede viajar la información desde las retinas, que están en perfecto estado, y otras áreas del cerebro. Se han llevado a cabo numerosas pruebas con sujetos que padecen de visión ciega, y ya no cabe duda de que su capacidad de adivinar diversas formas simples, la dirección del movimiento o la presencia o la ausencia de luz no es atribuible al azar (incluso hasta con el cien por cien de aciertos bajo ciertas condiciones). Ninguna persona con visión ciega ha mostrado aún la capacidad de discriminar colores dentro del campo ciego, aunque investigaciones recientes llevadas a cabo por Stoerig y Cowey (1990) parecen demostrar que ello puede ser posible.

¿Qué ocurre cuando hay visión ciega? ¿Se trata acaso, como han sugerido algunos filósofos y psicólogos, de percepción visual sin conciencia, como la que podría mostrar un mero autómata? ¿Constituye quizás una prueba en contra (o, en todo caso, un serio problema) para las teorías funcionalistas de la mente, al mostrarnos un ejemplo en el que todas las *funciones* de la visión siguen estando presentes, pero de donde se ha drenado todo el jugo de la *conciencia*? De ningún modo. En su prisa por arrimar la sardina de la visión ciega a su ascua ideológica, en ocasiones los filósofos han pasado por alto algunos hechos elementales sobre el fenómeno de la visión ciega y de la situación experimental en que dichos hechos pueden ser observados.

Como las «anomalías temporales» que analizamos en los capítulos 5 y 6, los fenómenos de la visión ciega sólo aparecen cuando tratamos a los pacientes desde el punto de vista de la heterofenomenología. Los experimentos no se podrían llevar a cabo si los investigadores no pudieran dar instrucciones verbales a los sujetos (y estar seguros de que se les ha comprendido), y las respuestas de los sujetos no constituyeran evidencia de este fenómeno tan sorprendente *sólo* cuando son interpretadas como actos de habla. Todo esto es tan obvio, que debo detenerme para insistir en ello.

La interpretación de la visión ciega es controvertida en muchos sentidos, pero existe un consenso bastante generalizado en un punto: todo el mundo acepta que el sujeto con visión ciega recibe información sobre un acontecimiento del mundo a través de los ojos (ésta es la parte de la «visión»), a pesar de carecer de la experiencia visual consciente de dicho acontecimiento

(ésta es la parte «ciega»). Más brevemente, la visión ciega comporta (1) la recepción de información visual que es (2) inconsciente a pesar de todo. La prueba de (1) es fácil: el sujeto muestra un elevado número de aciertos en test que buscan confirmar la presencia de esta información. La prueba de (2) es más circunstancial: los sujetos niegan ser conscientes de ninguno de estos acontecimientos, y sus negativas expresadas oralmente se ven confirmadas, por una parte, por los datos neurológicos sobre la lesión cerebral y, por otra, por la coherencia de estas negativas. ¡Así que les creemos![2]

Ésta no es una cuestión trivial. Nótese que lo sorprendente de la visión ciega desaparecería inmediatamente si llegáramos a la conclusión de que los sujetos están fingiendo, es decir, fingiendo no ser conscientes. O, por poner un caso más próximo, compárese nuestra aceptación de las negativas de los sujetos con visión ciega con el escepticismo con que recibimos las mismas negativas por parte de personas a las que se ha diagnosticado «ceguera histérica». Algunas veces, personas cuyos ojos y cerebros se hallan en perfecto estado, al menos desde el punto de vista de un fisiólogo, afirman haberse quedado completamente ciegas, y dan fe de este hecho comportándose «exactamente igual que un persona ciega». Existen razones por las cuales una persona se puede sentir motivada para «quedarse» ciega: sea como castigo contra sí misma o contra alguna otra que ahora deberá cuidarla y sentir pena por ella, sea como un modo de negar algún recuerdo visual terrible, o sea como una especie de respuesta de pánico ante alguna enfermedad o debilidad, de modo que es una ceguera «psicosomática», si es que se puede llamar ceguera. ¿Son realmente ciegas estas personas? Podría ser. Después de todo, se podría argumentar, si el dolor psicosomático es dolor real, y las náuseas psicosomáticas son lo bastante reales como para provocar el vómito, ¿por qué la ceguera psicosomática no puede ser una ceguera real?

Las personas con ceguera histérica afirman estar ciegas, aunque, al igual que los sujetos con visión ciega, muestran de forma irrefutable que están *recibiendo información visual.* Por ejemplo, las personas con ceguera histérica presentan un bajísimo nivel de aciertos cuando se les pide que adivinen rasgos visibles de las cosas. Éste es un claro signo de que están utilizando información visual para guiar su comportamiento a fin de cometer un máximo de «errores». Las personas con ceguera histérica poseen una misteriosa habilidad por encontrar sillas con las que tropezar. Y sin embargo, a diferencia de los verdaderos fingidores, las personas con ceguera histérica, cuando dicen que no tienen experiencias visuales, son sinceras; realmente creen en lo que están diciendo. ¿No deberíamos creerles nosotros también? ¿Cómo deberíamos tratar los textos de estos dos grupos diferentes de sujetos cuando extrapolamos lo que hay en sus mundos heterofenomenológicos?

2. Nótese que los detalles de la lesión neurológica en sí mismos (sin las negativas) no probarían nada; es sólo al contrastar la lesión neurológica con los testimonios (creíbles) y la evidencia comportamental, que podemos construir hipótesis sobre qué partes del cerebro son esenciales para qué fenómenos conscientes.

Aquí es donde la política ultraconservadora de la heterofenomenología produce sus beneficios. Tanto los sujetos con visión ciega como las personas con ceguera histérica son aparentemente sinceros es sus declaraciones de no ser conscientes de lo que ocurre en su campo ciego. Así pues, sus mundos heterofenomenológicos son parecidos, cuando menos en todo lo referente a su presunto campo ciego. Y sin embargo hay una diferencia. Poseemos menos conocimientos sobre las bases neuroanatómicas de la ceguera histérica de los que poseemos sobre la visión ciega, y, sin embargo, intuitivamente, tendemos a ser mucho más escépticos ante sus negativas.[3] ¿Qué nos hace sospechar que las personas con ceguera histérica no son *realmente* ciegas, que incluso son hasta cierto punto conscientes de sus mundos visuales? Las circunstancias sospechosamente propicias de su ceguera nos hacen dudar, pero más allá de esta evidencia circunstancial hay una razón más simple: dudamos de su ceguera porque, *sin ninguna ayuda*, las personas con ceguera histérica a veces *utilizan* la información que reciben de sus ojos de un modo que los sujetos con visión ciega no lo hacen.

Existe un factor típico de las situaciones experimentales con pacientes con visión ciega que encaja tan bien en nuestros supuestos que apenas nadie se molesta en discutirlo (pero véase Marcel, 1988; van Gulick, 1989; Carruthers, 1989): los sujetos con visión ciega necesitan alguna ayuda o indicación para poder «adivinar» con tanta precisión. El investigador debe haber dado las instrucciones iniciales: «Cuando oiga un tono, intente adivinar» o «Cuando yo le toque la mano, emita una respuesta». Sin estas indicaciones, el sujeto es totalmente incapaz de responder.[4]

Podemos evaluar nuestro diagnóstico de la diferencia imaginando una variante. Supóngase que halláramos un presunto sujeto con visión ciega que no necesitara indicaciones: «espontáneamente» produciría «conjeturas» (con un elevado nivel de aciertos aunque sin llegar al máximo) cada vez que algo se le presentara dentro del presunto campo ciego. Lo conduciríamos a nuestro laboratorio y llevaríamos a cabo el típico test a fin de realizar un mapa del escotoma; el sujeto nos avisa cada vez que la luz en movimiento desaparece dentro de su campo ciego, igual que cualquier otro sujeto con visión ciega. Pero, al mismo tiempo y de forma espontánea, proferiría comentarios

3. Sin los exámenes del cerebro que muestran la lesión cortical, no cabe duda de que también habría mucho escepticismo sobre los escotomas de los sujetos con visión ciega. Véanse, por ejemplo, Campion, Latto y Smith (1983), y Weiskrantz (1988).

4. El filósofo Colin McGinn (1991) dice de un paciente con visión ciega imaginaria: «Desde el punto de vista comportamental, puede funcionar casi igual que una persona que puede ver; desde el punto de vista fenomenológico, se considera ciega» (pág. 111). Esto es falso; desde el punto de vista comportamental no puede funcionar como una persona que puede ver. McGinn abunda en esta afirmación suya tan sorprendente: «Además, adoptando una actitud inocente, ¿acaso los pacientes con visión ciega no *parecen* tener experiencias visuales cuando llevan a cabo esas discriminaciones tan increíbles? (...) No parecen personas comportándose como si no hubiera *nada* que experimentar» (pág. 112). Esto también es falso. De hecho parece que *no* están teniendo experiencias visuales, *porque necesitan indicaciones*. Si no las necesitaran, parecería que estuvieran teniendo experiencias visuales, y nosotros no creeríamos sus negativas.

del tipo, «esto no es más que una conjetura, pero, ¿acaba usted de hacer brillar una luz dentro de mi escotoma?»; y esto lo haría justo después de que nosotros efectivamente hubiéramos encendido una luz. Ello resultaría, cuando menos, sospechoso, y podemos explicar por qué.

En general, cuando los sujetos siguen las instrucciones de un experimento, este hecho se considera como la evidencia de que han sido capaces de acatar las instrucciones porque han *experimentado de forma consciente* los estímulos relevantes. Por este motivo, la siguiente instrucción se consideraría como carente de sentido:

> «Cada vez que usted sea consciente de que se ha encendido una luz, presione el botón de la izquierda; cada vez que se encienda la luz y usted no es consciente de que se haya encendido, presione el botón de la derecha».

¿Cómo podría un sujeto seguir una instrucción semejante? Le estaríamos pidiendo lo imposible: condicionar su conducta a acontecimientos que le son inaccesibles. Sería como decir «levante una mano cada vez que alguien le guiñe un ojo sin que usted lo sepa». Un investigador nunca creería necesario insertar el adverbio «conscientemente», como en

«Cada vez que oiga conscientemente un tono, intente adivinar»

ya que se suele suponer que uno no puede condicionar sus propios protocolos sobre experiencias inconscientes, aunque este tipo de cosas pueda llegar a producirse. A fin de adoptar un protocolo como

«Cada vez que ocurra x, haga y»

uno debe ser capaz de ser consciente de que x ha ocurrido.

Ésta es la concepción tradicional, aunque este edificio de obviedades tiene una grieta. ¿Acaso no hemos aprendido que muchas de nuestras conductas se rigen por condiciones que sólo detectamos de forma inconsciente? Considérense los protocolos que regulan nuestra temperatura corporal, los que ajustan nuestro metabolismo, los que almacenan y liberan energía, o los que activan nuestro sistema inmunológico; considérense protocolos tales como el parpadeo cada vez que algo se aproxima o entra en nuestros ojos, e incluso conductas públicas a gran escala tales como el caminar (sin caerse) o el esquivar cuando un objeto aparece de repente a nuestro alrededor. Todos estos «comportamientos» se controlan sin la ayuda de la conciencia, como el propio Descartes ya observó en su momento.

Parece, pues, que hay dos tipos de protocolos comportamentales: aquellos que están controlados por el pensamiento consciente, y aquellos que están controlados por procesos «ciegos y mecánicos», igual que los procesos que controlan un elevador automático. Si un elevador automático debe se-

guir el protocolo de no levantar más de mil quilos, necesitará algún tipo de escala interna capaz de detectar cuándo se ha sobrepasado este límite. No hay duda de que un elevador automático no es consciente, ni de que no detecta nada de forma consciente y que, por tanto, carece por completo de protocolos conscientes. Podría decirse, sin embargo, que sigue protocolos que dependen de los diversos estados del mundo que es capaz de detectar e, incluso, que puede ajustar dichos protocolos en función de otras situaciones que es capaz de detectar, y así sucesivamente. Puede tener por tanto protocolos, metaprotocolos y metametaprotocolos, todos ellos dependientes de diversas y complejas combinaciones de situaciones detectadas, y todo sin un ápice de conciencia. Cualquier cosa que pueda hacer un elevador en el terreno de la detección y del seguimiento de protocolos, un cuerpo y un cerebro humano también pueden hacerlo. Pueden seguir complejos protocolos inconscientes como los de un elevador.

¿Cuál es entonces la diferencia entre el seguimiento inconsciente de protocolos y el seguimiento consciente de protocolos? Cuando consideramos los protocolos que nuestros cuerpos siguen inconscientemente, gracias a unos detectores de condiciones «ciegos y mecánicos», resulta tentador decir que, *puesto* que se trata de protocolos inconscientes, no son tanto nuestros protocolos como los protocolos *de nuestro cuerpo. Nuestros* protocolos son (podríamos decir que por definición) nuestros protocolos conscientes; aquellos que nosotros formulamos de manera deliberada y consciente, con la oportunidad de reflexionar (conscientemente) sobre sus pros y sus contras, y la oportunidad de ajustarlos y corregirlos a medida que una situación se desarrolla en nuestra experiencia.

Así pues, parece que cuando se adopta un protocolo como resultado de una discusión verbal, o como respuesta a unas instrucciones verbales, se trata *ipso facto* de un protocolo consciente, que debe depender de los acontecimientos que experimentemos de forma consciente (Marcel, 1988). Lo que resulta contradictorio es la idea de que uno pueda discutir y después decidir seguir un protocolo inconsciente, dependiente de acontecimientos detectados inconscientemente. El estatuto de dicho protocolo *podría* cambiar. Con un poco de práctica, y un poco de olvido estratégicamente situado, podríamos empezar con un protocolo adoptado y seguido conscientemente, para ir moviéndonos gradualmente hacia un estado de seguimiento de un protocolo *in*consciente, detectando los aspectos relevantes sin ser conscientes de ello. Esto podría darse, pero sólo si el vínculo con la consideración verbal del protocolo llegara a romperse.

Esta posible transición puede verse mejor si procedemos a la inversa. ¿Podría un sujeto con visión ciega *llegar a ser* consciente de las experiencias visuales en el escotoma por inversión del proceso que acabamos de imaginar? Después de todo, un sujeto con visión ciega manifiestamente recibe y analiza la información visual que de un modo u otro utiliza en sus conjeturas. Poco después de haberse producido el estímulo, algo ocurre en el cerebro del sujeto que marca el comienzo del estado *informado*. Si un observa-

dor externo (el investigador, por ejemplo) pudiera llegar a reconocer dichos comienzos, éste podría, en principio, hacerle llegar la información al sujeto. De este modo el sujeto podría llegar a reconocer los comienzos como información «de segunda mano» a pesar de ser incapaz de reconocerlos «de forma directa». Y, entonces, ¿acaso no podría ser el sujeto capaz en principio de «eliminar al intermediario» y de llegar a reconocer, al igual que lo hace el investigador, los cambios en sus propias disposiciones? Al principio, se necesita algún tipo de equipo de autocontrol —quizás el mismo que utiliza el investigador—, pero ahora sería el sujeto mismo quien miraría o escucharía las señales de salida.[5]

En otras palabras, ¿no sería posible en principio «cerrar el bucle de realimentación» e instruir así al sujeto a seguir un protocolo de condicionamiento de su conducta sobre cambios que él no experimentó («directamente»)? Planteo la perspectiva de este entrenamiento a la visión ciega como si se tratara de un experimento mental, aunque podría convertirse fácilmente en un experimento real. Podríamos intentar instruir a un sujeto con visión ciega a reconocer cuándo «adivinar».

Los sujetos con visión ciega son volubles en cuanto a sus talentos y disposiciones se refiere; hay días que están mejor que otros; mejoran con la práctica, a pesar del hecho de que normalmente *no* reciben información de retorno por parte del investigador sobre cómo lo están haciendo durante el experimento (como excepción, véase Zihl, 1980, 1981). Existen numerosos motivos por lo cuales esto debe de ser así, el principal de los cuales es que cualquier situación experimental de este tipo se podría ver invalidada por insinuaciones no intencionadas o inadvertidas por parte del investigador, de modo que las interacciones entre sujeto y experimentador se reducen al mínimo y se llevan a cabo siempre bajo un escrupuloso control. Sin embargo, los sujetos se abastecen de las indicaciones o la ayuda que reciben del investigador y, gradualmente, se acostumbran a lo que en otras circunstancias sería la tan poco prometedora práctica de hacer cientos o miles de conjeturas sobre cuestiones de las que están convencidos que no tienen ninguna experiencia. (Imagine cómo se sentiría usted si le hicieran coger el listín telefónico y adivinar qué marca de coche tiene cada una de las personas que aparece en la lista, sin que le dijeran si había acertado o no. Su motivación ante esta actividad no duraría mucho, a menos que usted tuviera una seguridad creíble sobre sus aciertos, y sobre por qué merecía la pena seguir adelante.)

¿Qué ocurriría, entonces, si decidiéramos dejar de lado todos los demás objetivos científicos y nos concentráramos en ver hasta dónde podemos llegar en la instrucción de un individuo con visión ciega, utilizando cualquier técnica de «*feedback*» que parezca ser de utilidad? Supóngase que empezamos con un sujeto con visión ciega normal, que «adivina» cada vez que le

5. «Si pudiera oír su propia respuesta galvánica sobre la piel, estaría en mejor forma». (Larry Weiskrantz, hablando sobre uno de sus pacientes con visión ciega, ZIF, Bielefeld, mayo de 1990.)

damos indicaciones (lo que se denomina respuesta de elección forzada), y cuyo nivel de aciertos es muy alto (si no lo fuera, entonces el sujeto no padecería de visión ciega). Gracias al *feedback*, el nivel de aciertos acabaría por alcanzar su techo, de modo que si el nivel de aciertos queda fijado dentro de un margen de precisión suficientemente alto, el sujeto tendrá la impresión de haber desarrollado una destreza útil y fiable que merecería la pena explotar. Éste es de hecho el estado en que se encuentran algunos sujetos con visión ciega hoy en día.

Supongamos ahora que empezamos a pedirle al sujeto que lo intente sin indicación alguna —es decir, que «adivine cuando hay que adivinar», que adivine «cuando el espíritu le mueva a hacerlo»—, y supongamos también que el investigador le proporciona al sujeto un *feedback* inmediato. Hay dos resultados posibles:

1. El sujeto se libra al azar y no mejora su actuación. A pesar del hecho de que el sujeto recibe información mensurable sobre el comienzo del estímulo, no parece haber manera de que el sujeto descubra cuándo se produjo esa información, independientemente de las ayudas de *biofeedback** que le proporcionemos.
2. El sujeto llega a ser capaz de operar sin ayuda del investigador (o sin ayudas de *biofeedback* temporales), mostrando un nivel de aciertos no atribuible al azar.

Evidentemente, el resultado que obtendríamos en cada caso particular es algo que sólo podemos determinar de forma empírica, y ni siquiera voy a aventurar una conjetura sobre la plausibilidad de obtener un resultado de tipo 2. Quizá, en cada caso individual ningún sujeto llegue a ser capaz de «adivinar» correctamente cuándo tiene que adivinar. Nótese, no obstante, que *si* un resultado de tipo 2 llegara a producirse, sería razonable pedir al sujeto que adoptara protocolos que le exigieran basar su conducta en estímulos cuyo transcurso sólo puede adivinar. Tanto si es consciente de dichos estímulos como si no lo es, si la fiabilidad de sus «conjeturas» es alta, podría tratar tales estímulos del mismo modo que cualquier experiencia consciente. Podría así elegir protocolos, tanto si se basaran en el transcurso de dichos estímulos como en el transcurso de acontecimientos experimentados conscientemente.

¿Pero sería esto suficiente para hacer al sujeto consciente de los estímulos? ¿Cuál es su intuición? Siempre que he preguntado, he recibido las más

* El término «biofeedback» podría equivaler, en español, a «biorrealimentación», término con el cual se denota una serie de técnicas que consisten en desarrollar la capacidad de poder controlar de forma voluntaria, aunque sea temporalmente, cualquier sistema neuronal. Dichas técnicas utilizan un procedimiento muy parecido al descrito en el texto, es decir, un observador externo produce una señal cada vez que el sistema neuronal está activo, de modo que el sujeto se hace consciente de la actividad del sistema y puede, así, llegar a controlarlo. [N. del T.]

variadas respuestas. La psicología del sentido común tampoco ofrece un veredicto claro. Sin embargo, disponemos del testimonio de un sujeto con visión ciega que habla sobre sí mismo ante una circunstancia similar. DB, uno de los sujetos estudiados por Weiskrantz, padece hemianopsia derecha y muestra las capacidades típicas de la visión ciega, con un elevado nivel de aciertos cuando recibe ayuda. Por ejemplo, si una luz se desplaza lentamente, en sentido horizontal o vertical, por su escotoma y se le pide que adivine si el movimiento es horizontal o vertical, su nivel de aciertos es bastante bueno, aunque niega toda conciencia del desplazamiento. Sin embargo, si la luz se mueve con más rapidez, ésta se convierte en un autoindicio: DB puede producir voluntariamente un informe bastante preciso del desplazamiento e, incluso, imitar el movimiento con un gesto de su mano, tan pronto como éste se ha producido (Weiskrantz, 1988, 1989). Cuando se le pregunta, DB no duda en afirmar que, *efectivamente*, ha experimentado el desplazamiento, ¿cómo podría ser capaz de atestiguarlo, si no? (Otros sujetos con visión ciega también atestiguan la experiencia consciente de estímulos que se mueven con rapidez.) Puede que queramos reservarnos nuestra opinión, pero su respuesta no debe sorprendernos, si el análisis de Rosenthal del concepto ordinario de conciencia va por buen camino. DB no sólo llega a estar informado sobre el desplazamiento de la luz; *se da cuenta* de que ha llegado a estar informado; en términos de Rosenthal, tiene un pensamiento de segundo orden cuyo contenido es que acaba de tener un pensamiento de primer orden.

Otto, nuestro crítico, vuelve a la carga:

> ¡Pero esto no es más que otro juego de manos! Siempre hemos sabido que los sujetos con visión ciega eran conscientes de sus conjeturas. Y todo esto no demuestra más que un sujeto podría desarrollar un talento para adivinar cuándo hay que adivinar (y, evidentemente, sería consciente de *aquellas* conjeturas). Llegar a reconocer que las conjeturas de uno sobre estos asuntos son fiables apenas sería suficiente para que uno se hiciera directamente consciente del acontecimiento sobre el que está conjeturando.

Esto parece indicar que se necesita algo más para la conciencia visual. ¿Qué se podría añadir? Por una parte, la conexión entre la conjetura y el estado que trata sobre ella, aunque fiable, parece bastante débil y efímera. ¿Podría hacerse más fuerte y duradera? ¿Cuál podría ser el resultado si los lazos de *concernencia* entre la conjetura y su objeto se multiplicaran?

3. Esconde el dedal: un ejercicio para despertar la conciencia

El término filosófico estándar para denominar la concernencia es *intencionalidad* y, según Elizabeth Anscombe (1965) se acuñó «por metáfora» a partir del latín, *intendere arcum in*, que significa *apuntar con un arco y una flecha a* (algo). Esta imagen del apuntar o de direccionalidad es central en

la mayoría de las discusiones filosóficas sobre la intencionalidad, aunque por lo general los filósofos han sustituido el complejo proceso de apuntar una flecha real por una mera flecha «lógica», una relación fundacional o primitiva, convertida en algo muy misterioso por su supuesta simplicidad. ¿Cómo *puede* algo dentro de nuestra cabeza apuntar esta flecha abstracta hacia una cosa en el mundo?[6] Puede que finalmente resulte correcto pensar en la relación de denotabilidad como una relación abstracta y lógica, pero al principio desvía nuestra atención de los procesos que realmente participan en el mantenimiento de la mente en contacto con las cosas del mundo, de modo que *efectivamente* se piense en ellas: los procesos de atender, mantener el contacto, seguir y rastrear (Selfridge, no publicado). El verdadero proceso de apuntar hacia algo, de «mantenerlo dentro del punto de mira», comporta llevar a cabo una serie de ajustes y compensaciones a lo largo del tiempo, bajo un «control de realimentación». Éste es el motivo por el cual la presencia de nubes de distractores (como las cintas metálicas que confunden a los sistemas de detección de misiles) puede hacer imposible la acción de apuntar. Fijarse en un objetivo el suficiente tiempo como para identificarlo es un logro que requiere algo más que una única y breve transacción de información. La mejor manera de mantener el contacto con algo consiste, literalmente, en mantener el *contacto* con ello, agarrarlo y ya no soltarlo, de modo que uno pueda examinarlo con tanto cuidado como quiera, tomándose su tiempo. La otra manera de hacerlo consiste en mantener el contacto de forma figurada, siguiéndolo con los ojos (y con el resto del cuerpo), sin perderlo nunca de vista. Éste es un logro que se puede conseguir con la percepción, evidentemente, pero no con una percepción pasiva; puede exigir algún esfuerzo, alguna planificación, y, en cualquier caso, una *actividad continuada* para mantener el contacto con algo.

Cuando era un niño me encantaba jugar a un juego que se llamaba Esconde el dedal. Se muestra a los participantes un sencillo dedal y todos, menos uno, que «esconde» el dedal, abandonan la habitación. Las reglas para esconder el dedal son claras: éste debe esconderse *bien a la vista*. No se puede colocar ni detrás ni debajo de nada, o en un lugar demasiado alto para ser visto por unos niños pequeños. Por lo general, en una sala de estar hay docenas de sitios donde se puede poner un dedal de modo que éste se integre perfectamente en su entorno, como un animal bien camuflado. Una vez está escondido, el resto de los niños vuelve a la habitación para encontrar el dedal. Cada vez que un niño lo localiza debe sentarse en un rincón de la habitación, tratando de no descubrir el lugar dónde está. Podemos contar con que los últimos niños que queden por encontrar el dedal lo habrán *mirado* más de una vez sin llegar nunca a verlo. En estos momentos tan emocionantes, todos pueden ver que el dedal está justo delante de las narices de Betsy, pongamos por caso, bien iluminado y subtendiendo un buen ángulo dentro de su campo visual. (En estos momentos, mi madre gustaba de de-

6. Mi respuesta a esta pregunta es mi libro *The Intentional Stance* (1987).

cir «¡Si fuese un oso, te mordería!») Las risas y las muecas de los demás niños harán pensar a Betsy que debe de estar mirando en la dirección correcta, y aun así no puede verlo.

Podríamos explicarlo de esta manera: aun si algún estado representacional en el cerebro de Betsy «incluye» en cierto modo el dedal, no hay todavía ningún estado perceptivo de Betsy que *concierna* al dedal. Podemos aceptar que uno de sus estados conscientes concierne al dedal: su «imágen de búsqueda». Puede que se esté concentrando intensamente en encontrar el dedal, el mismo dedal que había examinado sólo unos minutos antes, pero no se ha establecido todavía ninguna relación fuerte de intencionalidad o concernencia entre alguno de sus estados perceptivos y el dedal, aunque es muy posible que haya información en algún estado de su sistema visual que posibilitaría que alguien (un observador externo, por ejemplo, estudiando los estados visuales de su córtex visual) localizara o identificara el dedal. Para identificar el dedal, Betsy tiene que «concentrarse» en él, tiene que separarlo como «figura» del «fondo». Una vez se ha producido esto, Betsy realmente puede ver el dedal. Éste estará finalmente «en su experiencia consciente», y ahora que ya es consciente de su localización, Betsy podrá levantar su mano triunfalmente o ir a sentarse tranquilamente con los demás niños que ya habían localizado el dedal.[7]

Estos vínculos intencionados, guiados por *feedback*, limpios de errores y ajustados a las ganancias, son el requisito fundamental para un tipo de conocimiento que sea merecedor de este nombre, que pueda servir como base de un protocolo, por ejemplo. Una vez uno ha visto algo así, en este sentido fuerte del término, puede «hacer algo con ello» o hacer algo *porque* lo vio o *tan pronto como* lo vio. Una vez ha sido identificado, es fácil seguirle la pista a un dedal (a menos que uno se halle en una habitación llena de dedales durante un terremoto). En circunstancias normales, pues, el alto nivel que adquiere el dedal en el sistema de control de Betsy no es algo que dure unos breves instantes solamente; el dedal permanecerá localizado para Betsy durante el tiempo que ella tarde en alcanzarlo, o durante el tiempo necesario para confirmar su identidad, para verificarla de nuevo (una y otra vez, si es que queda alguna duda). Las cosas de las cuales somos más claramente conscientes son aquellos objetos que observamos con franqueza y sin precipitación, recogiendo e integrando el fruto de muchos movimientos sacádicos, construyendo un conocimiento durante el tiempo que mantenemos el objeto localizado en nuestro espacio personal. Si el objeto, como una mariposa, no para de moverse, deberemos tomar medidas para inmovilizarlo; así «podremos mirarlo», y si está bien camuflado en su entorno, deberemos dar algún paso —literalmente, si es que no podemos tocarlo— a fin de situarlo sobre un fondo que contraste.

7. ¿Es la identificación del dedal un efecto posterior a la conciencia, o una causa anterior a ella? Éste es el tipo de pregunta —¿orwelliano o estaliniano?— que el modelo de las Versiones Múltiples no enseña a no hacer.

Si no conseguimos hacer nada de esto, puede que no seamos capaces de ver el objeto, en un sentido importante y familiar del término.[8]

Es normal que los observadores de pájaros guarden una lista de las especies que han podido ver a lo largo de su vida. Suponga que usted y yo somos observadores de pájaros y que ambos oímos cantar a un pájaro en los árboles de nuestro alrededor; yo miro y digo, «lo ves, ¿no?». Usted mira en la misma dirección que yo y contesta, sinceramente, «no, no lo veo». Yo puedo añadir ese pájaro a mi lista, pero usted no, a pesar del hecho de tener la certeza moral de que su imagen debe de haberse paseado repetidamente por sus fóveas.

¿Qué diría usted? ¿Estaba el dedal «presente» en la conciencia de Betsy antes de que ella lo detectara? ¿Estaba el pájaro presente en el «fondo» de su conciencia, o no lo estaba? Pasar algo a un primer plano de la conciencia consiste en situarlo en una posición que nos permite emitir un testimonio sobre ello, pero ¿qué se necesita para que algo esté en un segundo plano de nuestra conciencia (y no solamente en un segundo plano de nuestro entorno visual)? El dedal y el pájaro estaban, sin lugar a dudas, dentro del entorno visual; no es éste el problema. Presumiblemente, no basta con que la luz reflejada por el objeto penetre en nuestros ojos, pero ¿qué otro efecto puede tener la luz reflejada —qué más puede anotar el cerebro sobre ello—, a fin de que el objeto pase de una situación favorable para una mera respuesta inconsciente al segundo plano de la experiencia consciente?

La única manera de responder a estas perplejidades «del punto de vista de la primera persona» es ignorar el punto de vista de la primera persona y ver qué podemos aprender desde el punto de vista de la tercera persona. En

8. En condiciones normales, la localización («detectarlo») y la identificación van muy unidas; detectar el objeto que queremos identificar es una condición previa para su identificación. Sin embargo, esta coincidencia oculta un hecho sorprendente: los mecanismo de identificación y los mecanismos de detección son, en gran medida, independientes en el cerebro, situados en regiones diferentes del córtex (Mishkin, Ungerleider y Macko, 1983), y es por tanto posible desactivarlos de forma independiente. Existen algunas patologías bastante raras en las que el sujeto puede identificar fácilmente qué está viendo sin poder localizarlo dentro de su espacio personal, y unas patologías complementarias en las que los sujetos pueden localizar un estímulo visual —pueden señalar en la dirección correcta, por ejemplo— y, sin embargo, apenas son capaces de identificar el objeto, a pesar de que, por lo demás, su visión es perfectamente normal. La psicóloga Anne Treisman (1988; Treisman y Gelade, 1980; Treisman y Sato, 1990; Treisman y Souther, 1985) ha llevado a cabo una serie de importantes experimentos que evidencian su hipótesis de que *ver* debe distinguirse de *identificar*. Cuando se ve algo, en su modelo, el cerebro elabora una «muestra» del objeto. Estas muestras son «representaciones episódicas temporales separadas», cuya creación es el necesario preámbulo de la identificación, algo que se consigue realizando una búsqueda en la memoria semántica utilizando un proceso parecido al que modelizan los sistemas de producción. Sin embargo, y si he comprendido bien el modelo de Treisman, una muestra no tiene por qué ir acompañada por una localización definitiva en el espacio personal, y, precisamente por este motivo, no sería de extrañar que aquellos sujetos que se encuentren en el mismo estado que Betsy (antes de encontrar el dedal) mostraran un elevado nivel de aciertos si se les pidiera que hicieran una conjetura de elección forzada sobre si el dedal se hallaba dentro de su campo de visión o no. Para algunos experimentos relacionados con este asunto, véanse Pollatsek, Rayner y Henderson (1990).

los capítulos 8-10, estudiamos un modelo de la producción de actos de habla que dependía de un proceso en pandemónium en el que la eventual unión de contenidos con expresiones constituía la culminación de una serie de competiciones, la construcción, el desmantelamiento y la reconstrucción de coaliciones. Los contenidos que participaron en la refriega, pero no consiguieron perpetuar su existencia por mucho tiempo, podrían enviar una especie de efecto de golpe «balístico» caracoleando por el sistema, pero seguirían siendo casi *inenarrables*. Cuando un acontecimiento no permanece, cualquier intento de narrarlo, si llega a iniciarse, o será abortado o vagará sin control, sin hallar nada con que contrastarse que le sirva para corregirse. Para que haya narrabilidad, tiene que haber una capacidad de identificar y de reidentificar el efecto. Podemos observar el desarrollo de la narrabilidad en muchos tipos de instrucción, próximos al tipo de instrucción que imaginamos para el sujeto con visión ciega: los resultados de entrenar el paladar en los catadores de vinos, de entrenar el oído en los músicos y cosas por el estilo, o el simple experimento de pulsar una cuerda de guitarra que describimos en el capítulo 3.

Considérense, por ejemplo, las instrucciones que se le darían a los aprendices de afinador de pianos. Se les dice que escuchen los «latidos» cuando pulsan la tecla que están afinando con referencia a una clave. *¿Qué* latidos? Al principio, la mayoría de novatos son incapaces de discernir dentro de sus experiencias auditivas nada que responda a la descripción de «latidos»; lo que oyen es algo que describirían más bien como una especie de sonidos disonantes o desafinados. Finalmente, sin embargo, si el proceso de instrucción se lleva a cabo con éxito, consiguen aislar, dentro de sus experiencias auditivas, los «latidos» de interferencia, y aprenden a percibir cómo cambian los patrones de ritmo como respuesta al ajuste del «martillo» de la clavija. Así, pueden afinar fácilmente el piano reglando esos latidos. En su relato de nuestra experiencia suelen decir —y nosotros podemos confirmarlo con episodios similares de nuestra propia experiencia— que como resultado de su instrucción, *su experiencia consciente ha cambiado*. En otras palabras, ha aumentado: ahora son conscientes de cosas de las cuales no eran conscientes antes.

En cierto modo, es evidente, siempre oyeron los latidos. Después de todo, eran conscientes de la interferencia que provoca el desafinamiento, pero al principio no eran capaces de detectar estos componentes en su experiencia, lo cual explica por qué podríamos decir que dichos factores contribuyeron a la experiencia, aunque no estaban *presentes en* ella. Desde el punto de vista funcional, el estatuto de dichas contribuciones antes de la instrucción es equivalente a los acontecimientos que se producen en la visión ciega: el sujeto es incapaz de referir dichas contribuciones, o de basar un protocolo en su comienzo, aunque el resultado de tales contribuciones puede manifestarse en la conducta del sujeto, por ejemplo en su capacidad de responder a preguntas planteadas con habilidad. Lo que quiero sugerir con esto es que estar en el segundo plano de la experiencia no tiene más secretos. Como he-

mos visto, no es descabellado suponer que un vínculo reforzado como el que acabamos de describir para el caso de los afinadores de piano y los catadores de vino pudiera establecerse también en el caso de un sujeto con visión ciega, hasta el punto de que éste declarara —declaración que nosotros no dudaríamos en aceptar— que se había *hecho* consciente de estímulos, aun en el primer plano de su conciencia, cuya existencia antes sólo podía conjeturar.

> ¡Un momento, no tan rápido! [nos dice Otto]. Tengo otra objeción: usted imagina al sujeto con visión ciega aprendiendo a utilizar sus capacidades de visión ciega a partir de estas nuevas técnicas, lo cual puede que le dote de una *especie* de conciencia de los acontecimientos que se producen en su campo ciego, pero esto sigue sin tener en cuenta algo. Esta conciencia sería conciencia *visual*; no sería como *ver*. Faltarían las «cualidades fenoménicas» o *qualia* de la visión consciente, incluso si el sujeto con visión ciega puede llevar a cabo estas operaciones funcionales.

Puede que sí, puede que no. ¿Qué son las «cualidades fenoménicas o *qualia*? (*Qualia* es la palabra latina que significa cualidades; su singular es *quale*.) En un principio nos parecen algo obvio —son la manera como nos aparecen, nos huelen, nos suenan, nos resultan al tacto las cosas—, pero tienen la característica de cambiar o de desaparecer cuando los sometemos a un examen más detallado. En el próximo capítulo nos dedicaremos a seguir a estos sospechosos a través de la espesura filosófica, pero antes debemos dedicar un poco más de tiempo a esas propiedades que no son cualidades fenoménicas, pero que podrían confundirse con ellas.

4. Visión protésica: ¿qué falta, además de la información?

¿Puede *ver* el movimiento DB, el sujeto de Weiskrantz? Bueno, lo cierto es que seguro que ni lo *oye* ni lo *siente*. ¿Es eso visión? ¿Posee las «cualidades fenoménicas» de la visión? Weiskrantz escribe:

> A medida que el estímulo se va haciendo cada vez más sobresaliente, el paciente podría insistir en que todavía no «ve», pero ahora posee una especie de «sensación» de que ahí hay algo. En algunos casos, si la prominencia del estímulo se aumenta aún más, se puede alcanzar un punto en que el sujeto «ve» pero la experiencia no es verídica. Por ejemplo, DB «ve» en respuesta a un estímulo que se mueve vigorosamente, pero no lo ve como un objeto coherentemente en movimiento, aunque refiere complejos patrones de «ondas». Otros sujetos refieren «sombras oscuras» que emergen a medida que el brillo y el contraste se sitúan en niveles altos (1988, pág. 189).

DB nunca percibe el objeto en movimiento como algo que tiene forma o color, pero ¿qué percibe entonces? Como probamos en el capítulo 2, en el experimento con un naipe situado en nuestra visión periférica, podemos

ver el naipe sin ser capaces de identificar ni su color ni su forma. Esto es visión normal, no visión ciega, de modo que, de acuerdo con ello, nos resistimos a negar que el sujeto tenga experiencia visual.

Figura 11.3

Un ciego con un sistema eléctrico portátil de 16 líneas. La cámara de televisión está conectada a un objetivo montado sobre la montura de las gafas. Una serie de cables lleva a un circuito de estímulos eléctricos (que el sujeto sostiene en su mano derecha). La matriz de 256 electrodos de plata concéntricos está en la mano izquierda del sujeto.

La pregunta de si esta manera anormal de obtener información acerca de las cosas visibles es una variedad de *visión* puede plantearse de manera más clara si desviamos nuestra atención hacia un caso que se aleja aún más de la visión normal. Se han diseñado dispositivos protésicos para permitir la «visión» a los ciegos, y algunos de estos dispositivos nos serán útiles para analizar los aspectos que nos interesan. Hace unos veinte años, Paul Bach-y-Rita (1972) desarrolló una serie de dispositivos que incluía unas pequeñas cámaras de vídeo de muy baja resolución que podían colocarse sobre la montura de unas gafas. La señal de baja resolución proveniente de estas cámaras, una matriz de 16-por-16 o de 20-por-20 *píxels*, se transmitía a la espalda

o el vientre del sujeto a través de una parrilla de bastoncillos vibrantes, eléctrica o mecánicamente, denominados tactores.

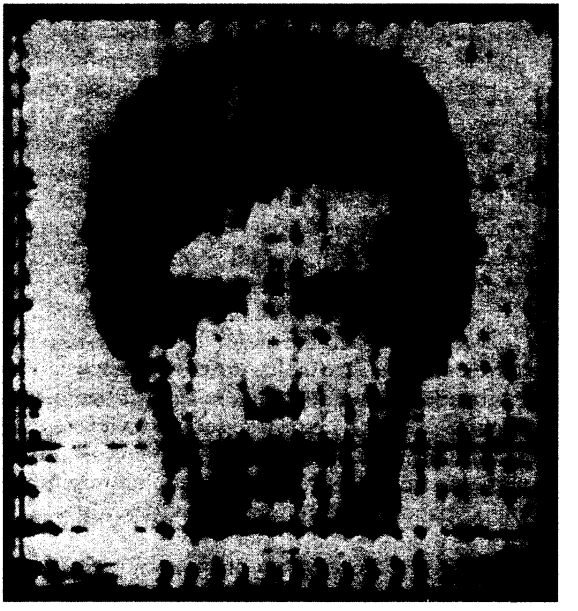


Figura 11.4

Apariencia de la representación, con una resolución de 400 puntos, de la cara de una mujer, en el monitor de un osciloscopio. Los sujetos pueden identificar correctamente patrones de estímulos a este nivel de complejidad.

Después de unas pocas horas de entrenamiento, los sujetos ciegos equipados con este dispositivo aprenden a interpretar patrones de vibraciones sobre su piel, igual que nosotros podemos interpretar las letras que alguien trace con un dedo sobre nuestra piel. La resolución es muy baja, pero aun así, los sujetos pueden aprender a leer signos, identificar objetos e incluso las caras de las personas, como podemos comprobar si *miramos* esta fotografía que muestra la señal tal como aparece en el monitor de un osciloscopio.

El resultado es, ciertamente, una experiencia perceptiva consciente producida protésicamente, pero dado que el estímulo se transmitía sobre la espalda o el vientre del sujeto en vez de su retina, ¿se trata de *visión*? ¿Posee las «cualidades fenoménicas» de la visión, o sólo de las sensaciones táctiles?

Recordemos uno de nuestros experimentos del capítulo 3. Desde el pun-

to de vista táctil, nos resulta fácil extender dichas sensaciones hasta la punta de un lápiz y sentir, así, texturas con ella, mientras que las vibraciones del propio lápiz sobre la piel de nuestros dedos nos pasan desapercibidas. No debe pues sorprendernos descubrir que los sujetos de Bach-y-Rita disfrutaran de un efecto similar, aunque más extremo. Después de un breve período de entrenamiento, la conciencia de la presencia de las vibraciones sobre su piel desaparecía; podríamos decir que la matriz de *píxels* se hacía transparente, y el punto de vista de los sujetos pasaba a ser el punto de vista de la cámara montada sobre sus cabezas. Una demostración sorprendente de la fuerza de este cambio de punto de vista es la conducta de un sujeto experimentado cuya cámara poseía una lente con *zoom*, controlado con un botón (págs. 98-99). La matriz de vibradores estaba colocada sobre su espalda y la cámara estaba montada a un lado de su cabeza. Cuando el experimentador, sin previo aviso, accionó el botón del *zoom*, provocando así que la imagen en la *espalda* del sujeto se expandiera o surgiera repentinamente, el sujeto se apartó instintivamente *hacia atrás, levantando sus brazos para proteger su cabeza*. Otra sorprendente demostración de la transparencia de las vibraciones la constituye el hecho de que los sujetos que habían sido entrenados con la matriz de vibradores en la espalda se adaptaban casi inmediatamente cuando dicha matriz pasaba a colocarse sobre su vientre (pág. 33). Y sin embargo, como señala Bach-y-Rita, seguían respondiendo a un picor en la espalda como algo que había que rascarse —no se quejaban porque lo «veían»— y eran perfectamente capaces de responder a las vibraciones como tales cuando se lo requerían.

Todas estas observaciones resultan ciertamente tentadoras pero inconcluyentes. Podríamos argumentar que una vez el uso de los dispositivos se convirtió en su segunda naturaleza, los sujetos estaban viendo realmente, o, por el contrario, que sólo alguno de los rasgos «funcionales» más centrales de la visión había sido reproducido de forma protésica. ¿Qué decir de las demás «cualidades fenoménicas» de la visión? Bach-y-Rita nos relata el resultado de haber mostrado a dos sujetos entrenados, estudiantes universitarios del sexo masculino, por primera vez en su vida, fotografías de mujeres desnudas de la revista *Playboy*. Se sintieron defraudados: «Aunque ambos eran capaces de describir con cierta fidelidad el contenido de las fotografías, la experiencia no mostró ningún componente afectivo; no despertó ninguna sensación de placer en ellos. La experiencia molestó mucho a los dos jóvenes, que sabían del componente afectivo que contenían fotografías similares para sus amigos dotados de un sentido normal de la visión» (pág. 145).

Así pues, los dispositivos protésicos de Bach-y-Rita no producían *todos* los efectos de la visión normal. Parte de estas deficiencias tienen que deberse a la notable diferencia en el nivel de flujo de la información. La visión normal nos informa de las propiedades espaciales de las cosas de nuestro alrededor a gran velocidad y casi con todo el nivel de detalle que deseemos. No debe sorprendernos que la información espacial de baja resolución en-

viada hacia el cerebro a través de una interficie en la piel no fuera capaz de provocar las reacciones que se provocan en personas con visión normal cuando sus sistemas visuales se ven inundados con información de entrada.[9] ¿Qué placer sería de esperar que sintiera un persona con visión normal si *mirara* traducciones de una baja resolución parecida —échele una ojeada a la figura 11.4— de fotografías de personas hermosas?

No está claro tampoco qué cambiaría si consiguiéramos mejorar el «nivel de baudios»[10] de la visión protésica para acercarlo al de la visión normal. Pudiera ser que con simplemente aumentar la cantidad y el nivel de información, proporcionando así mapas de *bits* de mayor resolución para el cerebro, fuese suficiente para producir el encanto que faltaba. O quizás algo de ese encanto. Las personas que son ciegas de nacimiento se encontrarían entonces en gran desventaja frente a los que han perdido la visión en algún momento de su vida, ya que no poseen ninguna de las asociaciones específicamente visuales que, sin duda, juegan un papel importante en el placer que sienten las personas videntes con sus experiencias, que les *recuerdan* experiencias visuales anteriores. También podría ser que parte de ese placer que sentimos con las experiencias visuales no fuese más que un efecto secundario de viejas huellas fósiles de una economía anterior en nuestros sistemas nerviosos, un asunto que ya tratamos en el capítulo 7 y que retomaremos en el capíulo siguiente.

Las mismas consideraciones son aplicables a la visión ciega y a cualquier mejora que podamos imaginar en las capacidades de los sujetos con visión ciega. Toda discusión de la visión ciega muestra una cierta tendencia a ignorar cuán insignificante es la información que los sujetos con visión ciega recogen en sus campos ciegos. Una cosa es ser capaz de adivinar, cuando se es requerido a hacerlo, si dentro del campo ciego apareció un cuadrado o un círculo. Otra cosa muy diferente sería ser capaz de adivinar con detalle, y cuando se es requerido a hacerlo, qué se ve a través de la ventana.

9. Por ejemplo, la latencia de respuesta para algunas de estas tareas perceptivas, incluso con sujetos muy entrenados, es bastante alta: de ocho a quince segundos para diversas identificaciones simples, por ejemplo (Bach-y-Rita, pág. 103). Ello demuestra por sí mismo que el flujo de la información en la visión protésica es, comparado con la visión normal, bastante lento.

10. El «nivel de baudios» es el término estándar para medir el flujo de información digital (significa, aproximadamente, *bits* por segundo). Por ejemplo, si su ordenador se comunica con otros ordenadores a través de la línea telefónica, puede transmitir cadenas de *bits* a 1.200 baudios o 2.400 baudios o incluso a niveles superiores. Se necesita un nivel de baudios aproximadamente cuatro veces más rápido para transmitir en tiempo real imágenes animadas en alta resolución —la situación típica en que una imagen vale más que mil palabras. Las señales de televisión son analógicas, como las de una grabación fonográfica, y no digitales, como las de un disco compacto, de modo que su flujo de información se mide en ancho de banda, en vez de en *nivel de baudios*. El término es anterior a la era de los ordenadores; el código Baudot, así llamado por su inventor (igual que el código Morse), fue el código telegráfico internacional estándar adoptado en 1880, y el nivel de baudios era el número de unidades de código transmitidas por segundo. Al utilizar el término «nivel de baudios» en vez de «ancho de banda» no pretendo sugerir que el tratamiento de la información llevado a cabo por el cerebro deba concebirse en términos digitales.

Podemos utilizar lo que hemos aprendido sobre la visión protésica para guiar nuestra imaginación hacia lo que sentiría un sujeto con visión ciega al recuperar más *funciones* de la visión. Intentemos imaginar una persona corticalmente ciega que, después de un arduo entrenamiento, (1) ha convertido su capacidad de adivinar cuándo hay que adivinar en su segunda naturaleza, (2) puede jugar a Esconder el dedal como el mejor, y (3) ha sido capaz de aumentar la rapidez y el detalle de sus conjeturas de forma notable. Nos lo encontramos leyendo el periódico y riéndose con los cómics y le rogamos que no explique lo que hace. He aquí tres situaciones hipotéticas en orden ascendente de plausibilidad:

1. «¡Meras conjeturas, claro! No veo ni torta, ¿sabe?, pero he aprendido a adivinar cuándo tengo que adivinar, y ahora, por ejemplo, adivino que usted acaba de hacer un gesto un poco feo y que en su cara hay una mueca de total incredulidad».
2. «Bueno, lo que empezó no siendo más que meras conjeturas, gradualmente fue perdiendo su carácter de conjetura, a medida que yo iba creyendo en ellas. Se convirtieron en, si se me permite usar la palabra, *presentimientos*. De repente empecé a *saber* que algo estaba ocurriendo en mi campo ciego. Así pude expresar mi conocimiento y actuar de acuerdo con ese conocimiento. Por otra parte, después empecé a tener un metaconocimiento de que yo era efectivamente capaz de tener tales presentimientos, y pude utilizar dicho metaconocimiento al planear mis acciones y al establecer mis propios protocolos. Lo que empezó como conjeturas conscientes se convirtió en presentimientos conscientes, que ahora se producen con tanta rapidez y tanta fuerza que apenas puedo separarlos. ¡Y, sin embargo, sigo sin ver ni torta! En todo caso no de la manera que veía antes. No es lo mismo que ver.»
3. «Bueno, pues, la verdad es que es muy parecido a ver. Ahora me muevo por el mundo sin esfuerzo gracias a la información sobre mi entorno recogida por mis ojos. O también, puedo ser consciente de lo que recibo de mis ojos si así lo deseo. Sin dudarlo, reacciono ante el color de las cosas, su forma y su localización, y ya he perdido toda la sensación de esfuerzo que experimentaba mientras desarrollaba este talento y lo convertía en mi segunda naturaleza.»

Aun así, podemos imaginar a nuestro sujeto afirmando que falta algo:

> «Los *qualia*. Mis estados perceptivos tienen qualia, por supuesto, pues son estados conscientes, pero antes de que perdiera la vista, eran qualia *visuales*, y ahora no lo son, a pesar de mi entrenamiento».

Puede que a usted esto le parezca obvio, que tiene sentido, que es precisamente lo que sería de esperar que dijera nuestro sujeto. Si es así, lo que queda de capítulo es para usted un ejercicio especialmente diseñado para

sacudir esta convicción. Si usted ya está empezando a dudar de que toda esta charla sobre qualia tenga algún sentido, es probable que haya anticipado algunos de los derroteros que va a tomar nuestra historia.

5. «Repleción» frente a averiguación

> Pero por el hecho de que haya este sentimiento de lo extraño no podemos decir: todo objeto que conocemos bien y que no nos es extraño nos da un sentimiento de confianza. Es como si quisiéramos decir que el lugar que una vez tomó el sentimiento de lo extraño debería estar ocupado de *alguna manera*.
> LUDWIG WITTGENSTEIN, *Investigaciones filosóficas*, 1953, §596*

En el capítulo 2, vimos que uno de los motivos para creer en el dualismo era su promesa de explicarnos cuál es la «sustancia de que están hechos los sueños»: las vacas de color violeta y otros producos de nuestra imaginación. En el capítulo 5 vimos las confusiones que suscita el supuesto, natural pero erróneo, de que *después* de que el cerebro ha efectuado una discriminación o un juicio, éste re-*presenta* el material sobre el que se basa el juicio, para el solo disfrute de la audiencia del Teatro Cartesiano, añadiendo los colores. Esta idea de *repleción** es común en el pensamiento de los más sofisticados teóricos, y no es otra cosa que un vestigio más del materialismo cartesiano. Lo divertido es que aquellos que más usan el término son los que más saben, pero ya que encuentran irresistible el término, se cubren las espaldas escribiéndolo entre comillas.

Por ejemplo, casi todo el mundo describe al cerebro «rellenando» el punto ciego (la cursiva en todos los ejemplos es mía):

> ...el conocido fenómeno neurológico de la «repleción» subjetiva de la parte ausente de un área ciega en el campo visual (Libet, 1985b, pág. 567).

> ...uno puede localizar su propio punto ciego, y también demostrar de qué modo un patrón se «*rellena*» o se «completa» en ese punto ciego (Hundert, 1987, pág. 427).

También hay «repleción» auditiva. Cuando escuchamos un discurso, los huecos en la señal acústica se pueden «rellenar», como ocurre, por ejemplo, en el «efecto de reconstrucción de los fonemas» (Warren, 1970). Ray Jackendoff lo expresa así:

* Edición bilingüe con traducción castellana de Alfonso García Suárez y Ulises Moulines, Instituto de Investigaciones Filosóficas (Universidad Nacional Autónoma de México) y Editorial Crítica, México y Barcelona, 1986 y 1988. [N. del T.]

* Tomo el término «repleción» del artículo «El problema de la consciencia» de Francis Crick y Christof Koch, publicado en el volumen monográfico de Investigación y Ciencia titulado *Mente y Cerebro*. [N. del T.]

> Considérese, por ejemplo, la percepción del habla con información de entrada defectuosa o llena de ruido (por ejemplo, en presencia de un avión a reacción con los motores encendidos o durante una conversación telefónica con problemas en la línea). (...) Lo que uno construye... no es solamente el significado pretendido por el emisor, sino también la estructura fonológica: uno «oye» más de lo que la señal transmite realmente. (...) En otras palabras, se «*rellena*» la información fonética a partir de estructuras de nivel superior y a partir de la señal acústica; y, a pesar de que existe una diferencia en cuanto a cómo ésta se deriva, no hay ninguna diferencia cualitativa en la estructura completa (Jackendoff, 1989, pág. 99).

Y, cuando leemos un texto, ocurre algo similar (pero visual); en palabras de Bernard Baars:

> Hallamos fenómenos parecidos en el conocido «efecto del lector de pruebas», aquella situación en que es difícil detectar los errores tipográficos en unas galeradas porque la mente «rellena» la información correcta (Baars, 1988, pág. 173).

Howard Margolis añade un comentario incontestable a este asunto de la «repleción»:

> Los detalles que se «*rellenan*» suelen ser correctos (Margolis, 1987, pág. 41).

El reconocimiento tácito de que hay algo que no funciona en esta idea de la «repleción» se manifiesta claramente en esta descripción del punto ciego realizada por el filósofo C. L. Hardin en su libro *Color for Philosophers*:

> Cubre un área con un diámetro visual de 6 grados, lo suficiente como para albergar la imagen de diez lunas llenas colocadas una junto a la otra y, sin embargo, no hay ningún hueco en la región correspondiente del campo visual. Ello se debe al hecho de que el ojo del cerebro *rellena* con cualquier cosa que esté a la vista en la regiones circundantes. Si es azul, *rellena* en azul; si es a cuadros, no percibimos ninguna discontinuidad en la expansión de los cuadros (1988, pág. 22).

Hardin no acaba de atreverse a decir que el cerebro rellena con los cuadros, ya que ello comporta, sin lugar a dudas, una sofisticada tarea de «construcción», como el «remiendo invisible» para tapar el siete en su americana de espiga por el que usted pagaría lo que fuese: las rayas coinciden y los tonos de color de la tela nueva no contrastan con la vieja. Parece como si rellenar de azul fuera una cosa —lo único que se necesita es una o dos pinceladas de pintura cerebral con el color adecuado—; pero rellenar a cuadros ya es otra cosa, algo que Hardin no se atreve a afirmar.

Sin embargo, como el propio Hardin nos hace notar, somos tan inconscientes de nuestros puntos ciegos cuando estamos ante un plano pintado a cuadros que cuando estamos ante un plano coloreado uniformemente, así

que, sea lo que sea lo que provoca esa inconsciencia, nuestro cerebro es capaz de llevarlo a cabo con la misma facilidad en ambos casos. «No percibimos ninguna discontinuidad», nos dice. Pero, si el cerebro no tiene que tomarse la molestia de rellenar los cuadros, ¿por qué tendría que tomársela para rellenar en azul?

Presumiblemente, en ninguno de los dos casos la «repleción» es verdadera repleción, como la que requeriría el recurso a algo parecido a unos pinceles. (Ésta es la moraleja de nuestra historia del sistema CADBLIND Mark II en el capítulo 8.) Presumo que nadie piensa realmente que la «repleción» sea un proceso en que el cerebro se dedica a cubrir un espacio con un *pigmento*. Sabemos que la imagen real invertida de la retina es el último estadio de la visión en que las cosas tienen color de la misma manera que lo tienen en la imagen de una película proyectada en una pantalla. Toda vez que no podemos hablar literalmente de la existencia de un ojo de la mente, tampoco podemos suponer que el cerebro necesite pigmentos.

Nada más que decir sobre pigmentos. Y a pesar de todo, tenemos la tendencia a pensar que algo ocurre en el cerebro que es de *alguna* manera muy importante, análogo a cubrir un área con un pigmento; en caso contrario no hablaríamos de «repleción». Es este acontecimiento tan especial, sea lo que sea, lo que, aparentemente, se produce en el «medio» especial de la experiencia visual o auditiva. Como dice Jackendoff, al hablar del caso concreto de la audición, «uno [oye] más de lo que la señal transmite realmente», pero nótese que Jackendoff escribe «oír» entre comillas. ¿Qué puede ser eso que está *presente* cuando uno «oye» sonidos que rellenan silencios o «ve» colores que cubren espacios vacíos? Parece que, efectivamente, hay algo en todos estos casos, algo que el cerebro tiene que proporcionar (para «rellenar»). ¿Cómo podríamos llamar a esta cosa desconocida? Llamémosle *figmento*.* La tentación, pues, consiste en suponer que hay algo, hecho de figmento que está ahí cuando el cerebro «rellena» y que no está cuando el cerebro no se preocupa de «rellenar». Dicho así, la idea del figmento no resultará atractiva para muchos. (Al menos eso espero.) Sabemos algo más: el figmento no existe. El cerebro no fabrica figmento; ni lo utiliza para rellenar huecos; el figmento no es más que una quimera, un producto de mi imaginación. ¡Nada más que decir sobre el figmento! Pero, entonces, ¿qué significa «repleción»? *¿Qué puede* significar, si no significa repleción con figmento? Si no existe un medio como el figmento, ¿en qué difiere la «repleción» de no tomarse la molestia de llevarla a cabo?

En el capítulo 10 vimos de qué modo representan los colores los sistemas de CAD, asociando un número con cada *píxel* o con cada región del objeto representado, y también vimos de qué modo el sistema CADBLIND Mark II podía buscar, o detectar, colores leyendo dicho código. Este proceso es muy parecido a un viejo pasatiempo para niños, los cuadernos para

* El autor hace un juego de palabras intraducible entre «pigment» ("pigmento") y «figment» ("quimera"). [N. del T.]

colorear, que nos proporcionan una simple analogía del proceso de representación que se produce, o que podría producirse, en el cerebro. La figura 11.5 es una representación que contiene información sobre formas, pero no sobre colores.

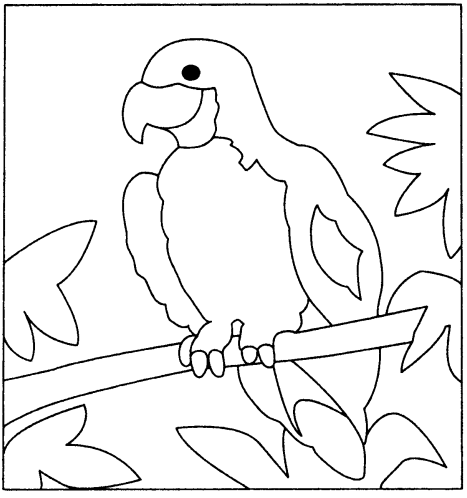


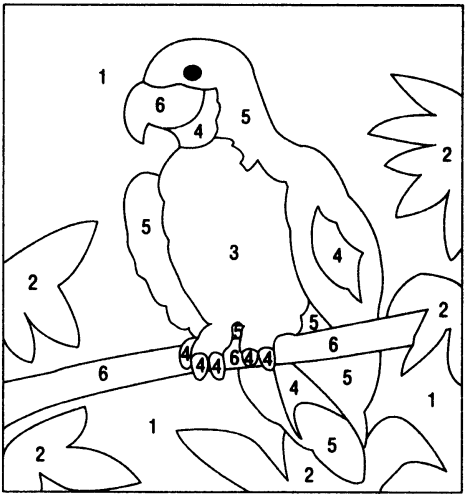
Figura 11.5

Compárese esta figura con la figura 11.6, que contiene información sobre colores en forma de código de colores. Si usted tomara unos lápices de colores y siguiera las instrucciones para pintar el dibujo, convertiría la figura 11.6 en un nuevo tipo de representación «rellenada»; una representación en que las regiones se han rellenado con colores reales, con pigmentos reales.

Existe otra manera de «rellenar» de color *píxel* a *píxel* mediante un mapa de *bits* codificado, como en la figura 11.7.

Las figuras 11.6 y 11.7 son, ambas, posibles formas de repleción (por comparación con la figura 11.5, por ejemplo), ya que cualquier procedimiento que necesite saber cuál es el color de una región determinada puede, a través de una inspección mecánica de esa región, extraer la información deseada. Esto es pura repleción informacional. Los sistemas son, claro está, totalmente arbitrarios. Sin ninguna dificultad podríamos construir infinidad de sistemas de representación con la misma funcionalidad, con diversos sistemas de codificación o con diferentes medios.

Si usted hace un dibujo en colores en su ordenador personal con un programa para PC como «Paintbrush», lo que ve en la pantalla se representa dentro de la máquina como un mapa de *bits* en el «*buffer* de fotogramas»,



1- Azul
2- Verde
3- Naranja
4- Rojo
5- Violeta
6- Amarillo

Figura 11.6

```
1111111111111111111111111111111111111111111111111111
1111111111111111111111111111111111111111111111111111
1111111111111111111555511111111111111111111111111111
1111111111111111155555555511111111111111111111111111
1111111111111115555●555555111111111111111111111111111
1111111111111166666666445555511111111111111111222222
1111111111111666666664455555111111111111111122222
1111111111116666666644555551111111111111112222222
1111111111116611144444555555111111111122222222222
1111111111111111111333333555555551111111111222222
1111111111111111333333335555555551111112222222222
1111111111111555333333333335555555511111112222222
1111111111115553333333333335555455551111122221122
1111111222211555533333333333355544555511111111122
1111222222211155553333333333335544445551111111122
1222222221111155553333333333335544444551111111111
2222222211111115553333333333335544444551111111111
2222222111111115555333333333335555445551122222222
2222222111111115555333333333335555555511222222222
2222222111111115555333333333335555555566666622222
2221122211111111111533335333333666666666666622212
1222111111111111664333366434666666666661112211211
1111111166666666666444446644464455555555111211112
1666666666666661111111111155554445555555111111111
6666666661111111111111111111555544444555551111111111
6666111221111111111111111111554455544551111111111
1111222222111111111111111111111545555555122221111
1222222211111111111122222221112255555552222222111
2222222222221111111112222222212222555552222222211
2222222222222221111111122222222222222222221111111
2222211222221111111111111122222222222222221111111
```

Figura 11.7

de forma parecida a lo que tenemos en la figura 11.7, pero cuando usted guarda el dibujo en un disco, un algoritmo de compresión lo traduce a algo parecido a la figura 11.6. El algoritmo divide el área en regiones de igual coloración y almacena el perímetro de cada región y su número de color en una carpeta de «archivo».[11] Una carpeta de archivo tiene una fidelidad en el almacenamiento de datos equiparable al mapa de *bits*, pero generaliza la información concerniente a cada región, etiquetando cada una de ellas sólo una vez; es un sistema de representación más eficiente.

Un mapa de *bits*, al etiquetar cada *píxel* de forma explícita, constituye una forma de lo que podríamos denominar representación *toscamente continua*, donde el grado de tosquedad está en función del tamaño de los *píxels*. No puede decirse que un mapa de *bits* sea realmente una imagen, sino más bien una matriz de valores, una especie de receta para formar una imagen. La matriz puede almacenarse en cualquier sistema que conserve la información sobre localizaciones. El vídeo es otro medio de representación toscamente continua, pero lo que se almacena en la cinta no son imágenes, sino recetas (con diferentes grados de precisión) para formar imágenes.

Otra manera de almacenar la imagen en la pantalla de su ordenador podría ser la de tomar una fotografía en color y almacenar así la imagen, por ejemplo, en una diapositiva de 35mm, y este sistema difiere de los demás de un modo importante: aquí sí que hay tintes que cubren regiones en un espacio real. Como en el mapa de *bits*, ésta es una representación toscamente continua de las regiones espaciales dibujadas (continua hasta el nivel del grano de la película; al nivel de la finura de grano, la representación ya no es continua sino granular, como las representaciones por *píxels*). Pero a diferencia del mapa de *bits*, se utilizan colores para representar los colores. Un negativo en color también utiliza colores para representar los colores, pero en un espectro invertido.

Aquí tenemos, pues, tres maneras diferentes de «rellenar» información sobre colores: el color vía un código de números, como en la figura 11.6 o en una carpeta de archivo, color por mapa de *bits*, como en la figura 11.7, en un *buffer* de fotogramas o una cinta de vídeo, y color por colores. El sistema del código de números es, desde un determinado punto de vista, una manera de «rellenar» información sobre colores, pero posee una mayor eficiencia, en comparación con las demás técnicas, precisamente porque no recurre a una repleción real y explícita de cada *píxel*. ¿En cuál de estos tres sentidos (en el supuesto de que exista alguno) podemos decir que «rellena» el cerebro un punto ciego? Nadie cree que el cerebro utilice números almacenados en registros para codificar los colores, pero esto no es más que un ardid para desviar nuestra atención del asunto principal. La idea de números almacenados en registros puede ser simplemente un modo de referirse a un sistema de magnitudes, cualquier sistema de «vectores» que el cerebro

11. Existen algoritmos de compresión que no se basan en la divisón del dibujo en regiones del mismo color, pero no voy a ocuparme de ellos aquí.

pudiera utilizar como «código» de colores; podría tratarse de frecuencias de actividad neuronal, o de algún sistema de direcciones o localizaciones en redes neuronales, o cualquier otro sistema de variaciones físicas en el cerebro que a usted se le pudiera ocurrir. Los números almacenados en registros poseen la útil propiedad de preservar las relaciones entre magnitudes físicas al tiempo que son neutrales en cuanto a las propiedades «intrínsecas» de dichas magnitudes, de modo que pueden corresponderse con cualquier magnitud física que el cerebro utilice para «codificar» los colores. Aunque los números se pueden utilizar de formas completamente arbitrarias, también pueden utilizarse de maneras no arbitrarias, para reflejar las relaciones estructurales entre colores que se han descubierto. El conocido «sólido de colores», donde el tono, la saturación y el brillo son las tres dimensiones a lo largo de las cuales nuestros colores varían,[12] es un espacio lógico ideal para un tratamiento numérico, cualquier tratamiento numérico que refleje las relaciones entre elementos, las relaciones de oposición y complementariedad, etc., que muestra la visión humana. Cuanto más aprendamos sobre el sistema que utiliza el cerebro para codificar los colores, más poderoso y menos arbitrario será el modelo de la visión humana en color que seremos capaces de construir.

El problema de hablar del cerebro como un «codificador» de colores que utiliza intensidades o magnitudes de una cosa u otra es que puede inducir al incauto a pensar que eventualmente estos códigos tienen que descodificarse, lo cual nos lleva «de vuelta a los colores». Ésta es una ruta —quizá la más popular— que nos devuelve al problema del figmento: uno imagina que el cerebro podría almacenar inconscientemente su información enciclopédica sobre el color en un formato parecido al de la figura 11.8, pero se las arregla para «descodificar» una representación en forma de «colores reales» para ocasiones especiales, como pasar una cinta de vídeo para proyectar color real en una pantalla. Ciertamente, existe una diferencia desde el punto de vista fenomenológico entre recordar la proposición de que la bandera es roja, blanca y azul e imaginar la bandera «en color» y «ver» (con el ojo de la mente) que es roja, blanca y azul. Si este contraste en la fenomenología induce a algunas personas a postular el figmento, aún más convincente resulta el caso del fenómeno de la repleción del color de neón (van Tuijl, 1975), del que tenemos un ejemplo en la contraportada de este libro.

El color rosa que usted ve rellenando el anillo delimitado por las líneas

12. Otras criaturas poseen sólidos de colores distintos, o hipersólidos. Nosotros somos «tricrómatas»: tenemos tres tipos de células traductoras fotopigmentadas en los conos de nuestras retinas. Otras especies, como las palomas, son tetracrómatas; su espacio de colores subjetivo se representaría, numéricamente, como un hiperespacio en cuatro dimensiones. Otras especies son dicrómatas, y todas sus discriminaciones de colores podrían proyectarse sobre un único plano bidimensional. (Nótese que «blanco y negro» es un esquema de representación unidimensional, con una escala de todos los grises posibles representable como distancias diferentes en una línea entre 0 y 1.) Para algunas reflexiones sobre las implicaciones de la inconmensurabilidad de los sistemas de colores, véanse Hardin (1988) y Thompson, Palacios y Varela (en prensa).

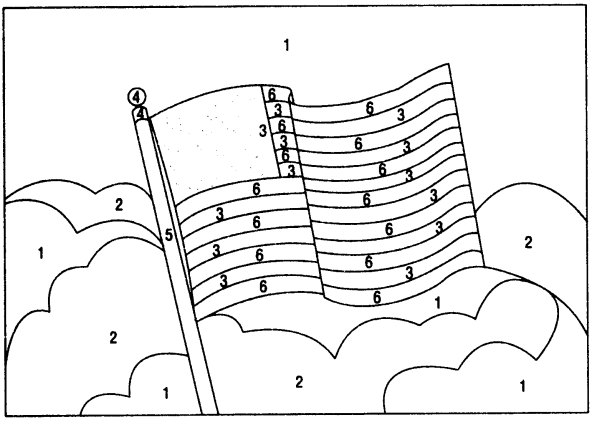


Figura 11.8

rojas no está causado por un corrimiento de la tinta en el papel, ni por una dispersión de la luz. Ese rosa no está en su imagen retinal, es decir, no hay más color que el de las líneas rojas. ¿Cómo podemos explicar esta ilusión? Un circuito cerebral, especializado en las formas, se ve inducido a error al distinguir una determinada región con unos límites definidos: el anillo con su «contorno subjetivo». Hay muchas figuras similares, como las siguientes, que inducen contornos subjetivos.

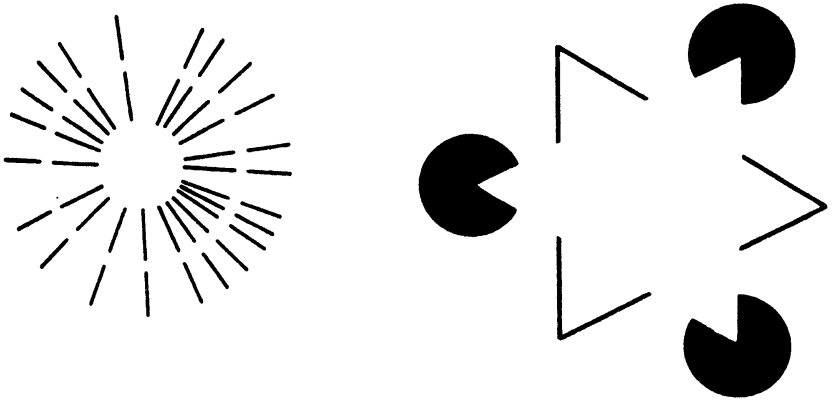


Figura 11.9

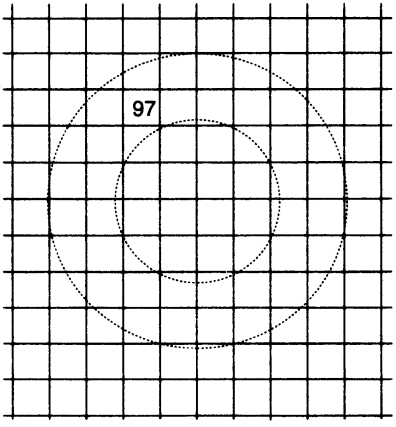


Figura 11.10

Otro circuito cerebral, esta vez especializado en colores, pero no muy capaz en todo lo que concierne a la forma y la localización en el espacio, llega a una discriminación de color (rosa #97, pongamos por caso) con la que «etiquetar» algo cercano, y la etiqueta se asigna (o «asocia») a toda la región.

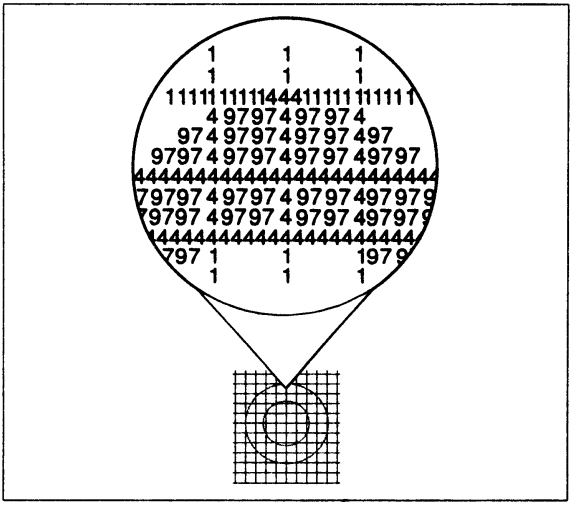


Figura 11.11

El hecho de que se produzcan dichas discriminaciones en condiciones como las descritas constituye todavía un aspecto controvertido, aunque la controversia se centra en los mecanismos causales que conducen a un etiquetado erróneo de la región, y no en los «productos» adicionales (si es que los hay) del sistema visual. Pero, ¿no falta algo todavía? No he profundizado mucho en mi explicación según la cual la región recibe una etiqueta dentro de un sistema de coloración por códigos numéricos: ¿acaso esta receta para colorear una imagen no debe ejecutarse en alguna parte? ¿Acaso no hay que «rellenar» con rosa #97? Después de todo, insistirá usted, ¡se ve de color rosa! En cualquier caso, no se ve una región bien delimitada y marcada con un número. El rosa que vemos no está en el mundo exterior (es pigmento o tinta o «luz coloreada», así que debe estar «aquí dentro»; en otras palabras, es figmento rosa.

Debemos ser muy cuidadosos en el momento de distinguir la hipótesis del «figmento rosa» de otras hipótesis que podrían ser alternativas legítimas a una explicación que no va más allá de la idea de un código numérico para los colores. Por ejemplo, podría darse el caso de que en algún lugar del cerebro hubiera una representación toscamente continua de las regiones coloreadas —un mapa de *bits*—, de modo que «cada *píxel*» de la región lleve la etiqueta «color #97», más o menos como se muestra en la figura 11.11.

Ésta es una posibilidad empírica. Podríamos desarrollar experimentos para confirmarla o desmentirla. La pregunta sería: ¿existe un medio representacional en el cerebro en el que el valor de algún parámetro variable (la intensidad o cualquier otra cosa que codifique el color) deba propagarse o replicarse por los *píxels* relevantes de una matriz, o, por el contrario, sólo hay una «única etiqueta» para la región, sin necesidad de que haya «repleción» o «expansión»? ¿Qué tipo de experimentos favorecerían un modelo así del fenómeno de la repleción del color de neón? Sería impresionante, por ejemplo, que se pudiera demostrar que bajo determinadas condiciones el color se expande lentamente, derramándose desde las líneas rojas hasta alcanzar las lindes subjetivas que delimitan el contorno de la figura.[13] No quiero prejuzgar la cuestión, ya que mi objetivo principal al plantearla era el de ilustrar mi afirmación de que aunque quedan muchas cuestiones por resolver sobre cómo se produce el fenómeno de la repleción del color de neón

13. Como si fuera a propósito para responder a esta sugerencia, V. S. Ramachandran y R. L. Gregory (presentado para publicación) acaban de llevar a cabo una serie de experimentos con lo que ellos denominan (equivocadamente, a mi modo de ver) escotomas artificialmente inducidos, según los cuales parece haber datos que evidencian un proceso de repleción gradual de las texturas y los detalles. Existe una diferencia fundamental entre sus circunstancias experimentales y las condiciones que he descrito; en sus experimentos hay una competencia entre dos fuentes de información, y una va perdiendo terreno (gradualmente). El fenómeno de la repleción espacial gradual de texturas es un importante descubrimiento, pero no nos conduce más allá de un modelo en la línea del de la figura 11.11. Por otra parte, quedan todavía una serie de aspectos por resolver sobre los mencionados experimentos antes de poder otorgarles una interpretación definitiva.

en el cerebro, ninguna de ellas comporta diferencias sobre si se genera figmento durante la «descodificación» del sistema de codificación neuronal.

La pregunta de si el cerebro «rellena» de una manera o de otra no es una pregunta que la introspección nos pueda ayudar a dilucidar, pues, como vimos en el capítulo 4, la introspección nos proporciona —tanto al sujeto como al investigador «exterior»— sólo el contenido de la representación, no los rasgos del medio representacional mismo. Para obtener datos sobre el medio, necesitamos efectuar nuevos experimentos.[14] Pero por lo que a algunos fenómenos respecta, ya podemos estar *bastante* seguros de que el medio de representación es una versión de algo eficiente, como el código numérico para los colores, y no toscamente continuo, como el mapa de *bits*.

Considérese cómo debe de operar el cerebro con el papel pintado, por ejemplo. Supongamos que usted entra en una habitación y observa que el papel pintado es una matriz regular de pequeños veleros idénticos, o —en homenaje a Andy Warhol— de retratos fotográficos idénticos de Marilyn Monroe. A fin de poder identificar un dibujo como un retrato de Marilyn Monroe, usted tiene que recorrer el dibujo con sus fóveas: la imagen tiene que caer bajo las fóveas de alta resolución de sus ojos. Como comprobamos en el experimento con el naipe del capítulo 2, su visión *parafoveal* (la que sirve el resto de la retina) no tiene muy buena resolución; ni siquiera puede identificar una J de corazones sostenida con el brazo extendido. Y sin embargo, sabemos que si entrara en una habitación cuyas paredes estuvieran forradas con fotos idénticas de Marilyn Monroe, usted reconocería «al instante» que la decoración era ésa. En una fracción de segundo usted vería que había «muchos y muchos retratos idénticos, detallados y bien enfocados de Marilyn Monroe». Toda vez que los movimientos sacádicos de sus ojos se producen con una frecuencia del orden de unas cuatro o cinco veces por segundo como máximo, usted puede recorrer con sus fóveas sólo una o dos Marilyns en el tiempo que necesita para llegar a la conclusión y *después ver* cientos de Marilyns idénticas. Sabemos que la visión parafoveal no podría distinguir una Marilyn de una mancha con forma de Marilyn y, sin embargo, lo que usted ve *no* es un papel pintado con una Marilyn en el centro rodeada por cientos de manchas con forma de Marilyn.

¿Es acaso posible que el cerebro tome una de sus visiones foveales de alta resolución de Marilyn y la reproduzca, como si hiciera fotocopias, por la proyección interna del plano de la pared? Éste es el único modo en que los detalles de alta resolución que usted utilizó para identificar a Marilyn podrían «pasar al segundo plano», ya que la visión parafoveal no es lo bastante aguda como para aportarlos por sí misma. Supongo que, en principio,

14. Por ejemplo, los primeros experimentos de Roger Shepard sobre la rotación mental de diagramas cúbicos demostraron que los sujetos parecían albergar representaciones toscamente continuas en rotación de las formas que estaban imaginando; se necesitaron nuevos experimentos, centrados en las propiedades temporales de las representaciones subyacentes, para confirmar (parcialmente) la hipótesis de que estaban haciendo lo que parecía que estaban haciendo. (Véase Shepard y Cooper, 1982.)

es posible, pero dudo que el cerebro se tome todas las molestias de llevar a cabo *esta* repleción. Habiendo identificado una Marilyn, y al no recibir información de que las demás manchas no son Marilyns, llega a la conclusión de que éstas son también Marilyns, y asocia a toda esa región una etiqueta que indica «más Marilyns», sin ocuparse de reproducir todas las Marilyns.[15]

Evidentemente, a usted no le parece que sea así. A usted le parece que está viendo cientos de Marilyns idénticas. Y en cierto modo, así es: sí que hay cientos de Marilyns idénticas en el papel pintado de la pared, y usted las está viendo. Lo que no es cierto, sin embargo, es que haya cientos de Marilyns idénticas representadas en su cerebro. Su cerebro sólo representa, de un modo u otro, que *hay* cientos de Marilyns idénticas, y por muy clara que sea su impresión de que las ve todas con detalle, el detalle está en el mundo, y no dentro de su cabeza. Y no se utiliza ningún figmento en la reproducción de esa apariencia, porque no se reproduce ninguna apariencia, ni siquiera como un mapa de *bits*.

Ahora podemos responder a nuestra pregunta sobre el punto ciego. El cerebro no tiene que «rellenar» el punto ciego, ya que la región donde cae el punto ciego ya está etiquetada (por ejemplo, «a cuadros» o con «Marilyns» o, simplemente, «más de lo mismo»). Si el cerebro recibiera datos contradictorios sobre alguna región, abandonaría o ajustaría su generalización, pero no recibir datos del punto ciego sobre la región no es lo mismo que recibir datos contradictorios. La ausencia de datos sobre la región procedentes del punto ciego no constituye ningún problema para el cerebro; dado que el cerebro no dispone de precedentes de recibir información de ese punto de la retina, no ha desarrollado ninguna agencia epistémicamente hambrienta esperando ser alimentada con información procedente de esa región. De todos los homúnculos de la visión, ninguno tiene el cometido de coordinar información proveniente de esa región del ojo, de modo que cuando no llega información de ese punto, nadie se queja. El área queda, simplemente, desatendida. En otras palabras, todas las personas con visión normal «sufren» de una leve «anosognosia». No somos conscientes del «déficit», del hecho de que no recibimos información visual de nuestros puntos ciegos. (Un excelente examen de la anosognosia es McGlynn y Schacter, 1989.)

El punto ciego es un vacío espacial, pero también puede haber vacíos temporales. Los más pequeños son aquellos que se producen cuando nuestros ojos efectúan sus movimientos sacádicos. No percibimos estos vacíos, pero tampoco es preciso que se rellenen *porque* estamos diseñados para no percibirlos. Los análogos temporales de los escotomas podrían ser las «ausencias» que se producen durante las crisis epilépticas. Estos vacíos son percibidos por quien sufrió el ataque, pero sólo por inferencia: no pueden «ver los bordes», al igual que usted no puede ver los bordes de sus puntos ciegos, pero se pueden ver sorprendidos, retrospectivamente, por discontinuidades en los acontecimientos que han experimentado.

15. En el Apéndice B, propondré algunos «experimentos con papel pintado» que podrían servir para verificar esta afirmación.

El principal fallo en la idea de «repleción» reside en el hecho de que comporta la idea de que el cerebro está aportando algo, cuando la verdad es que el cerebro está ignorando algo. Lo cual lleva a muchos grandes pensadores a cometer crasos errores perfectamente personificados por Edelman: «Una de las características más sorprendentes de la conciencia es su continuidad» (1989, pág. 119). Esto es totalmente falso. Una de las características más sorprendentes de la conciencia es su *dis*continuidad, como nos revelan hechos como el del punto ciego o los vacíos sacádicos, por recurrir a los ejemplos más simples. La discontinuidad de la conciencia es sorprendente precisamente por la *aparente* continuidad de la conciencia. Neumann (1990) señala que la conciencia puede, por lo general, ser un fenómeno lleno de vacíos y, en tanto que los límites temporales no son percibidos con claridad, no habrá ninguna sensación de vacío en el «flujo» de la conciencia. En palabras de Minsky, «[N]ada *parecerá* discontinuo, salvo que sea *representado* así. Paradójicamente, nuestra continuidad proviene, no de una genuina perceptividad, sino de nuestra maravillosa *insensibilidad* ante la mayor parte de los cambios» (1985, pág. 257).*

6. La negligencia como pérdida patológica del apetito epistémico

El lema del cerebro para tratar con los puntos ciegos podría ser: no me hagas preguntas y no te contaré mentiras. Como vimos en el capítulo 1, en tanto en cuanto el cerebro pueda saciar su apetito epistémico, éste no necesita nada más. ¿Pero qué ocurre en aquellas ocasiones en que hay mucho menos apetito epistémico del que debería haber? Éstas son las patologías de la desatención.

Una de las formas más comunes de desatención es la *hemidesatención*, en la que una parte del cuerpo, normalmente la izquierda, queda totalmente desatendida, debido a una lesión cerebral en el lado opuesto. Se desatiende en estos casos no sólo la mitad izquierda del cuerpo, sino también la parte izquierda de todo lo que está próximo. Si un grupo de personas se sitúa alrededor de la cama del paciente, éste sólo mirará hacia los que están a su derecha; si se le pide que cuente el número de personas que hay en su habitación, tenderá a ignorar a aquellas que están a su izquierda, y si alguien desde su izquierda intenta llamar su atención, no lo conseguirá. Sin embargo, se puede demostrar que los órganos sensoriales del paciente siguen recibiendo información y analizando y respondiendo a los estímulos que se producen a la izquierda. ¿Qué puede estar ocurriendo en la cabeza de estos pacientes? ¿Acaso la «parte izquierda del espacio fenoménico está en blanco»? ¿O el «ojo de la mente» del paciente no es capaz de ver el material que el cerebro le presenta en la parte izquierda del escenario del Teatro Cartesiano?

Existe una explicación más simple, no en términos de representaciones

* Página 265 de la edición castellana, (Galápago, 1986). [N. del T.]

internas con propiedades curiosas, sino en términos de *desatención*... ¡en el sentido político del término! Daniel Patrick Moynihan es conocido por sus recomendaciones de que ciertos problemas raciales de los Estados Unidos se resolverían por sí mismos si se los trataba con una «desatención benigna»: si Washington y el resto del Estado los ignoraran por un tiempo. No creo que ésta sea una buena recomendación, aunque Moynihan tenía su parte de razón: existen circunstancias en las que la desatención benigna es necesaria, como por ejemplo en nuestro tratamiento del problema del punto ciego.

Según lo hemos expresado aquí, no hay homúnculos que deban «ocuparse» de la información provinente de la parte del campo visual cubierto por el punto ciego, así que cuando no llega nada, tampoco hay nadie para quejarse de ello. Quizá la única diferencia entre nosotros y los que padecen de desatención patológica y otras formas de anosognosia es que algunos de sus homúnculos lastimeros han muerto. Esta teoría ha sido propuesta, en términos menos pintorescos, por el neuropsicólogo Marcel Kinsbourne (1980), que denomina a nuestros homúnculos lastimeros «analizadores corticales». En términos del modelo que hemos desarrollado, la desatención podría describirse como una pérdida de influencia política por parte de ciertos grupos de demonios del cerebro, debida, en la mayoría de los casos (aunque no en todos), a la muerte o a la eliminación de su representante. Dichos demonios siguen siendo activos, intentando llevar a cabo diversas tareas, e incluso con éxito en algunos casos, aunque carecen de la fuerza suficiente para ganar ciertas competiciones ante coaliciones mucho mejor organizadas.

En este modelo, la negligencia benigna de nuestros puntos ciegos se funde casi imperceptiblemente con las muchas otras desatenciones disfuncionales leves que padecemos, y después con las desatenciones más raras que estudian los neurólogos. Por ejemplo, yo mismo padezco diversas formas comunes de desatención. No muy grave, pero un tanto embarazosa en ocasiones, es mi desatención a los *errores tipográficos*. Soy patológicamente incapaz de percibir los errores tipográficos en las pruebas de imprenta de mis artículos cuando las releo, y sólo gracias a los más laboriosos ejercicios de concentración consigo superarlo. No es, como ha sugerido Baars, que mi cerebro «rellene» con las formas ortográficamente correctas, ya que por lo general no presta la atención suficiente a estos asuntos para percatarse de estos errores; su atención se ve captada por otros aspectos de las palabras que llenan la página. Otra de mis pequeñas incapacidades es la desatención a los *exámenes de mis alumnos*. Es increíble lo atractivas que me resultan tareas tales como fregar el suelo de la cocina, cambiar el papel que forra los estantes de mi armario o comprobar el estado de cuentas de mi talonario, cuando tengo sobre mi mesa un montón de exámenes por corregir. Esta característica, el elevado interés que se siente por cualquier alternativa, es particularmente evidente en la hemidesatención; como primera aproximación, cuanto más a la derecha está algo, más merecedor es de la atención de los pacientes con hemidesatención. Sin embargo, mi forma más seria de desa-

tención es mi grave *desatención financiera*. En efecto, tan poco me gusta comprobar el estado de cuentas de mi talonario, que sólo alguna altenativa realmente espantosa, como por ejemplo corregir exámenes, es capaz de llevar mi atención hacia el asunto. Esta desatención tiene serias consecuencias para mi bienestar, consecuencias que es fácil hacerme notar, pero a pesar de estos avisos, cuya falta de efectividad resulta alarmante, consigo persistir en mi desatención, a menos que me esfuerce por tomar medidas drásticas de automanipulación.

Y no es que *no pueda ver* mi talonario, es que *ni siquiera lo miro*. Y aunque en momentos fríos y reflexivos como éstos, soy capaz de explicar todo esto (lo cual prueba que no padezco de anosognosia *profunda* sobre mi propia discapacidad), en el curso normal de los acontecimientos, parte de lo que no percibo es mi propia desatención de mis finanzas. Anosognosia leve, pues. Desde este punto de vista, lo único sorprendente de las formas más raras de negligencia que estudian los neuropsicólogos es el tipo de cosas que son desatendidas. Imagine a alguien que desatiende todo lo que está *a su izquierda* (Bisiach y Luzzatti, 1978; Bisiach, 1988; Bisiach y Vallar, 1988; Calvanio, Petrone y Levine, 1987). O imagine a alguien que ha perdido la visión del color *pero no se lamenta de ello* (Geschwind y Fusillo, 1966). O incluso, imagine a alguien que se ha quedado ciego, pero todavía no se ha percatado de su profunda pérdida, el síndrome de Anton o la negación de la ceguera (Anton, 1899; McGlynn y Schacter, 1989, págs. 154-158).

Estas condiciones son fáciles de explicar dentro de la teoría de las Versiones Múltiples de la conciencia, donde el testigo central ha sido sustituido por coaliciones de especialistas cuyos apetitos epistémicos particulares no pueden ser adoptados inmediatamente por otros agentes, si los primeros han sido eliminados o se han tomado unas vacaciones.[16] Cuando estos apetitos epistémicos desaparecen, lo hacen sin dejar rastro, dejando el campo libre para otras coaliciones, otros agentes con otras agendas.

Pero el mismo principio que explica la desatención nos proporciona una alternativa para «los qualia visuales ausentes» de nuestro imaginario virtuoso con visión ciega. Mi sugerencia era que si se lamentaba de la ausencia de qualia, *podría* ser que percibiera la pobreza de la información que ahora recibe de su sistema visual, pero que su descripción fuese errónea. Especulé sobre la idea de que si pudiéramos aumentar el «nivel de baudios» de su dispositivo de recogida de información, algunos, si no todos, de los huecos entre su visión y la visión normal podrían cerrarse. Ahora vemos que otra manera más económica de cerrar el mismo hueco consistiría simplemente en rebajar sus niveles de apetito epistémico o en rebajar su curiosidad visual de alguna manera. Después de todo, si una persona con el síndrome de

16. Como comparación, puede consultarse Bisiach y otros (1986) y McGlynn y Schacter (1989), cuyos modelos de la anosognosia son similares, pero muy comprometidos con la «cajolología» de sistemas separados, especialmente el modelo de McGlynn y Schacter, quienes proponen un *sistema de conocimiento consciente* que recibe información de entrada proveniente de módulos.

Anton puede estar completamente ciega y no darse cuenta, un poco de desatención estratégicamente dosificada podría convertir a un sujeto con visión ciega que se lamenta de la pérdida de sus qualia visuales en un sujeto que no se queja y que asegura que ha recuperado la visión por completo. Parece ser que podríamos saber más, pero, ¿es cierto eso? ¿Le faltaría algo a una persona como ésta? En la visión normal no hay ningún *figmento*, así que no puede ser eso lo que falta. ¿Qué otra cosa podría ser?

7. Presencia virtual

> Experimentamos la sensación de realidad cuando toda pregunta formulada a nuestro sistema visual es respondida con tanta velocidad que parece como si esas respuestas ya hubieran estado allí.
> MARVIN MINSKY, *La sociedad de la mente*, 1985, pág. 257*

De nuevo, la ausencia de representación no es lo mismo que la representación de una ausencia. Y la representación de una presencia no es lo mismo que la presencia de una representación. Pero esto es difícil de creer. Nuestro convencimiento de que estamos *directamente informados* de las propiedades o rasgos especiales de nuestra experiencia es una de las intuiciones más poderosas con las que se enfrenta cualquiera que intente desarrollar una buena teoría de la conciencia. He estado usurpando su autoridad, intentando minarla, pero todavía me queda mucho trabajo por hacer. A Otto todavía le queda una vía que intentar:

> Lo que usted plantea con el problema de las Marilyns en el papel pintado, irónicamente, no constituye más que una defensa implícita del dualismo. Usted utiliza persuasivos argumentos en contra de la idea de que en el cerebro hay cientos de Marilyns de alta resolución, ¡y concluye que no están en ninguna parte! Mi argumento, por el contrario, consiste en afirmar que habida cuenta de que *lo que* yo veo son cientos de Marilyns de alta resolución, entonces, *ya que*, como usted demuestra, no están en ningún lugar de mi cerebro, deben estar en alguna otra parte: en mi mente no física.

Esos cientos de Marilyns del papel pintado parecen estar presentes en su experiencia, parecen estar en su mente, y no sólo en la pared. Como sabemos, sin embargo, toda vez que su mirada puede moverse en una fracción de segundo a fin de recoger información de cualquier parte de su entorno visual, ¿por qué motivo su cerebro debería molestarse en importar todas esas Marilyns? ¿Por qué no permitir que las almacene en el propio mundo, sin coste alguno, hasta el momento en que se las necesita?

Comparemos el cerebro con una biblioteca. Algunas bibliotecas de cen-

* Página 265 de la edición castellana (Galápago, 1986). [N. del T.]

tros de investigación son gigantescos almacenes que contienen millones de libros, todos ellos accesibles en sus estantes. Otras bibliotecas poseen menos libros, pero disponen de un generoso y eficiente sistema de compra de nuevos libros o de solicitud de préstamo a otras bibliotecas, utilizando un rápido sistema de préstamo interbibliotecario. Si no se guardan los libros en el edificio el acceso es más lento, pero no mucho más. Podríamos imaginar un sistema electrónico de préstamo interbibliotecario (vía fax o vía archivos informatizados) capaz de conseguir un libro en el mundo exterior en menos tiempo del que tardaría el más rápido de los corredores en traer el libro desde su estante. Un informático diría que, en un sistema como éste, los libros han estado siempre presentes en la biblioteca «de forma virtual», o que la «colección de libros virtuales» de la biblioteca es cien o mil veces mayor que la colección de libros de papel.

Ahora bien, ¿cómo es posible que nosotros sepamos, en tanto que usuarios de nuestras propias bibliotecas cerebrales, qué elementos de los que recuperamos siempre estuvieron *ahí*, y cuáles tuvo que ir a buscar nuestro cerebro, en rápidas incursiones en el mundo exterior para recoger información? Una serie de experimentos, cuidadosamente llevados a cabo de acuerdo con el método heterofenomenológico, podrían llegar a responder esta pregunta, aunque la introspección por sí sola es incapaz de distinguir. Esto no nos impide pensar, sin embargo, que somos capaces de distinguir. A falta de indicios que inclinen la balanza en una u otra dirección, tendemos naturalmente a concluir que hay más de lo que hay en realidad. He bautizado esta tendencia con el nombre de Trampa de la Introspección (Dennett, 1969, págs. 139-140), mientras Minsky la denomina Ilusión de la Inmanencia: «Siempre que somos capaces de responder a una pregunta sin una demora perceptible, parece como si esa respuesta ya estuviera activa en nuestra mente» (Minsky, 1985, pág. 155).*

La analogía del sistema de préstamo interbibliotecario es útil pero incompleta, ya que nuestro cerebro no posee la capacidad de adquirir información sobre cualquier cosa del mundo exterior que nos interese; también posee, literalmente, millones de centinelas que, casi sin interrupción, vigilan una porción del mundo exterior, prestos a dar la alarma y *llamar* nuestra atención sobre cualquier elemento novedoso y relevante que se produce en el mundo. En la visión, este cometido lo cumplen los bastoncillos y conos parafoveales de las retinas, y los agentes neuronales en una posición más interior que los centinelas, que están especializados en la detección del cambio y el movimiento. Si alguno de estos agentes da la alarma —«¡Cambios en mi sector!»—, casi instantáneamente se produce un movimiento sacádico, llevando la fóvea hacia la región de interés, de modo que la novedad puede ser localizada, identificada y tratada. El sistema de centinelas es tan fiable que resulta muy difícil introducir un cambio en el mundo visible sin que todo el sistema visual llegue a estar informado de ello, aunque con la ayuda

* Pág. 160 de la edición castellana (Galápago, 1986). [N. del T.]

de algunos trucos de alta tecnología, a veces es posible evitar a los centinelas, con resultados sorprendentes.

Cuando nuestros ojos se mueven en sacadas, las contracciones musculares que hacen rotar a los globos oculares son acciones *balísticas*: nuestros puntos de fijación son misiles no guiados cuyas trayectorias en el momento del despegue determinan dónde y cuándo alcanzarán su objetivo. Por ejemplo, si usted está leyendo un texto en la pantalla de su ordenador, sus ojos saltarán sobre algunas palabras en cada sacada, y saltarán más lejos y más deprisa cuanto mejor lector sea usted. ¿Qué ocurriría si un mago, una especie de espíritu maligno como el de Descartes pero a una escala menor, cambiara la palabra durante los pocos milisegundos en que sus ojos saltaban hacia ese nuevo destino? Aunque parezca mentira, un ordenador equipado con un rastreador de ojos automático puede detectar y analizar el despegue en los primeros milisegundos del movimiento sacádico, calcular cuál será el punto terminal del movimiento y, *antes de que éste haya concluido*, borrar la palabra en el punto terminal de la pantalla y sustituirla por otra palabra de la misma longitud. ¿Qué ve usted? Sólo la nueva palabra, y sin tener la sensación de que nada haya cambiado. Cuando usted examina el texto de la pantalla, tiene la sensación de que el mundo es tan estable como si las palabras del texto estuvieran grabadas en mármol, pero para otra persona que lea el misto texto detrás de sus espaldas (y con unos movimientos sacádicos procediendo a un ritmo distinto), la pantalla bulle de cambios.

El efecto es abrumador. La primera vez que me encontré ante un experimento con rastreadores de ojos, y comprobé hasta qué punto los sujetos permanecían totalmente ignorantes (aparentemente) de los cambios que se producían en la pantalla, pedí ser utilizado como sujeto. Quería comprobarlo por mí mismo. Me senté ante el aparato y me hicieron morder una barra, lo que permitía que mi cabeza permaneciera inmóvil. Esto facilita el trabajo del rastreador de ojos, que lanza un haz de luz imperceptible hacia la lente del ojo del sujeto, y analiza el retorno para detectar cualquier movimiento del ojo que se haya podido producir. Mientras esperaba que los investigadores pusieran en marcha el aparato, me dediqué a leer el texto de la pantalla. Esperé, esperé y esperé, ansioso porque empezara el experimento. Me impacienté. «¿Por qué no lo ponéis en marcha?», pregunté. «*Está* en marcha», me contestaron.

Dado que todos los cambios en la pantalla se producen durante los movimientos sacádicos, los centinelas no consiguen dar ninguna alarma efectiva. Hasta hace poco, el fenómeno se conocía con el nombre de «supresión sacádica». Se suponía que el cerebro debe, de un modo u otro, cerrar la entrada de información por los ojos durante los movimientos sacádicos, ya que nadie puede percibir los cambios que se producen en el campo visual durante dichos movimientos, y, evidentemente, nadie se queja por los repentinos y alarmantes cambios. Sin embargo, un inteligente experimento con un rastreador de ojos (Brooks y otros, 1980) ha demostrado que si el estímulo —como, por ejemplo, una palabra o una letra del alfabeto— se desplaza en

sincronía con las sacadas, siguiendo el paso de la «sombra» de la fóvea a medida que ésta se dirige hacia su nuevo punto de destino, éste es visto e identificado fácilmente por el sujeto. Así pues, no se bloquea la entrada de información desde el ojo hacia el cerebro durante los movimientos sacádicos, aunque bajo condiciones normales ésta no puede ser utilizada —todo va demasiado deprisa como para tener sentido—, de modo que el cerebro la trata con una desatención benigna. Si todos los centinelas se ponen a dar la alarma a la vez, lo mejor es ignorarlos a todos.

En la situación experimental en que participé, las palabras en la pantalla se borraban y se sustituían durante mis movimientos sacádicos. Si nuestra visión parafoveal no puede discriminar la palabra en el punto de partida antes de que el movimiento sacádico llegue hasta ella, una vez alcanzamos el punto en que se encuentra y la identificamos, no puede existir ningún registro o recuerdo previo en nuestro cerebro con que compararla. El cambio no se percibe porque la información lógicamente necesaria para percibirlo *no está ahí*. Sin duda, cuando leemos la página, *a nosotros nos parece* que todas las palabras de la línea están, en cierto sentido, presentes en nuestra conciencia (en un segundo plano), incluso antes de que les prestemos una atención específica, pero esto no es más que una ilusión. Sólo están presentes de forma virtual.

Evidentemente, en nuestro cerebro hay *alguna* información sobre las palabras cercanas, la suficiente como para haber servido de guía y de instigadora del siguiente movimiento sacádico, por ejemplo. ¿Qué información ya está *ahí* exactamente? Los experimentos con rastreadores de ojos y aparatos similares pueden determinar los límites de lo que podemos percibir, y determinar así los límites de lo que está *presente* en nuestra mente. (Véanse, por ejemplo, Pollatsek, Rayner y Collins, 1984; Morris, Rayner y Pollatsek, 1990.) Insistir, como Otto se siente tentado a hacer, en que lo que no está ahí en el cerebro debe no obstante estar *ahí* en la mente porque *parece* ciertamente estar ahí no tiene sentido. Pues, como hemos podido comprobar, no estaría «ahí» en ningún sentido que pudiera ser relevante *para las propias experiencias de Otto*, por no hablar de superar test, presionar botones, etc.

8. Ver es creer: un diálogo con Otto

Llegados a este punto, Otto, nuestro crítico, insiste en que hagamos un repaso, pues está seguro de que en algún punto del razonamiento lo hemos engañado. Le voy a hacer participar en un diálogo, con la esperanza de que haga justicia a la mayor parte de las dudas del lector, si no a todas. Otto empieza:

> «Me parece que usted ha negado la existencia de la mayoría de los fenómenos indudablemente reales que existen: las apariencias reales de las que no pudo dudar ni Descartes en sus *Meditaciones*».

En cierto sentido, está usted en lo cierto; eso es precisamente lo que estoy negando que exista. Volvamos por un momento al fenómeno de la repleción del color de neón. En la contraportada, parece haber un anillo rosado ligeramente brillante.

«No cabe duda de que es así.»

Pues no hay ningún anillo rosado y brillante. No realmente.

«De acuerdo. ¡Pero parece haberlo!»

Efectivamente.

«¿Dónde está, pues?»

¿Dónde está qué?

«El anillo rosado y brillante.»

No hay ninguno; creía que usted acababa de reconocerlo.

«Bueno, sí, no hay ningún anillo rosado ahí en la página, pero parece haber uno.»

Efectivamente, parecer haber un anillo rosado y brillante.

«Hablemos pues de *ese* anillo.»

¿Cuál?

«Ése que *parece* haber.»

No hay ningún anillo rosado que simplemente parezca haber.

«Mire usted, ¡yo no estoy diciendo que parezca haber un anillo rosado y brillante, sino que *realmente parece haber* un anillo rosado y brillante!»

No puedo más que estar de acuerdo. ¡Nunca podré acusarle de falta de ingenio al hablar! Eso es realmente lo que quiere decir cuando dice que parece haber un anillo rosado y brillante.

«Mire. No es sólo lo que quiero decir. No se trata sólo de que yo *piense* que parece haber un anillo rosado y brillante, ¡es que *realmente* parece haber un anillo rosado y brillante!»

Ya está. Ya ha caído en la trampa, como muchos otros. Parece que usted piensa que hay una diferencia entre pensar (juzgar, decidir, ser de la más sincera opinión de) que algo le parece a usted rosado y que algo *realmente le parece* a usted rosado. Pero no hay diferencia. No hay ningún fenómeno de apariencia real, además del fenómeno de juzgar de un modo o de otro que algo es algo.

Recuerde el papel pintado de Marilyn. La pared está, en efecto, cubierta de Marilyns de alta resolución. Además, eso es lo que a usted le parece. Le parece que la pared está cubierta de Marilyns de alta resolución. ¡Vaya suerte! Su aparato visual le ha permitido llegar a una creencia cierta sobre un rasgo de su entorno. Pero no hay muchas Marilyns *aparentemente reales* representadas en su cerebro, o en su mente. No hay ningún medio que *reproduzca* el detalle del papel pintado, que lo *represente* para su testigo interno. Todo lo que ocurre es que a usted le parece que ahí hay muchas Marilyns de alta resolución (y esta vez está usted en lo cierto, realmente las hay). En otras ocasiones usted puede no estar en lo cierto; le puede parecer —como en el fenómeno phi de los colores— que se movió un único punto de luz, cambiando de color a medida que se movía, cuando la verdad es que sólo había dos puntos de luz diferentes encendiéndose consecutivamente. El hecho de que a usted le parezca así no requiere ningún proceso de *reproducción* en el cerebro, del mismo modo que los juicios sobre colores del cerebro, una vez emitidos, no tienen por qué *descodificarse* después en alguna parte.

> «Pero, entonces, ¿qué está ocurriendo cuando me parece que hay un anillo rosado y brillante? ¿Qué explicación positiva puede darme su teoría? Se está mostrando usted muy evasivo sobre este asunto.»

Me temo que tiene usted razón. Ha llegado el momento de hablar claro y presentar una explicación positiva, pero le confieso que tendré que empezar con una caricatura que revisaré después. No parezco ser capaz de hallar una manera más directa de exponerlo.

«Ya me he dado cuenta. Adelante.»

Suponga que *hubiera* un Significador Central. Pero suponga que en vez de estar sentado en el Teatro Cartesiano contemplando la presentación, el Significador Central está sentado a oscuras y tiene presentimientos; de repente se le ocurre que podría haber algo rosado ahí fuera, del mismo modo que a usted le pareciera de repente que detrás suyo hay alguien.

«¿Qué son exactamente los presentimientos? ¿De qué están hechos?»

Buena pregunta, que de momento sólo puedo responder de forma evasiva, con una caricatura. Estos presentimientos son proposiciones que el Significador Central se exclama a sí mismo en su lenguaje privado, el menta-

lés. Así que su vida consiste en una secuencia de *juicios*, que son oraciones en mentalés, expresando una proposición después de otra, a una velocidad tremenda. Algunas de estas proposiciones, decide hacerlas públicas, en su traducción castellana.

Esta teoría tiene la virtud de eliminar el figmento, la proyección en el espacio fenoménico, la repleción de todos los espacios blancos en la pantalla del teatro, pero todavía tiene un Significador Central y el lenguaje del pensamiento. Revisemos, pues, la teoría. Suprimamos, primero, el Significador Central, distribuyendo todos sus juicios por el cerebro en el espacio y en el tiempo: todo acto de discriminación o de discernimiento o de fijación de contenido se produce en algún lugar, pero no hay un discernidor que haga todo el trabajo. En segundo lugar, suprimamos el lenguaje del pensamiento; el contenido de los juicios no tiene por qué ser expresable en forma «proposicional», eso es un error, un caso de proyección errónea y demasiado entusiasta de las categorías del lenguaje en las actividades del cerebro.

«Así que los presentimientos son como actos de habla pero sin actor y sin habla.»

Bueno, pues, sí. Lo que realmente hay es diversos acontecimientos de fijación de contenido produciéndose en diferentes lugares y tiempos en el cerebro. Ésos no son los actos de habla de nadie, y por tanto no tienen por qué estar en un lenguaje, pero son muy *parecidos* a actos de habla; tienen contenido, y tienen el efecto de informar a varios procesos con ese contenido. En los capítulos 5-10 consideramos algunas versiones más detalladas de esto. Algunas de estas fijaciones de contenido tienen efectos adicionales, que eventualmente conducen a la proferencia de enunciados —en un lenguaje natural— bien públicos o bien simplemente internos. Así se crea un texto heterofenomenológico. Cuando este texto se interpreta, se crea la ilusión benigna de que hay un autor. Esto es suficiente para producir *hetero*fenomenología.

«¿Pero, qué me dice de la fenomenología *real*?»

No existe. Recuerde nuestra discusión sobre la interpretación de ficciones. Cuando nos encontramos ante una novela que, aunque sea veladamene, podemos calificar de autobiografía, observamos que podemos proyectar los acontecimientos ficticios en muchos acontecimientos reales de la vida del autor, de modo que en un sentido amplio, podemos decir que la novela es *sobre* esos acontecimientos reales. Puede que el autor no se dé cuenta de ello, y sin embargo en este sentido amplio es cierto; es de esos acontecimientos que trata el texto, porque ésos son los acontecimientos reales que explican por qué *este* texto fue creado.

«¿Pero de qué trata el texto en un sentido no amplio?»

De nada. Es una ficción. *Parece tratar* de varios personajes, lugares y acontecimientos ficticios, pero dichos acontecimientos nunca se produjeron; no trata *realmente* de nada.

> «¡Pero cuando yo leo una novela, estos acontecimientos ficticios vuelven a la vida! A mí me ocurre algo; yo *visualizo* los acontecimientos. El acto de leer, e interpretar, un texto como una novela crea cosas nuevas en mi imaginación: imágenes de los personajes llevando a cabo sus actos. Después de todo, cuando vamos a ver la versión cinematográfica de una novela que hemos leído, a menudo pensamos, "¡No es así como yo me lo había imaginado!".»

De acuerdo. En «Fearing fictions» el filósofo Kendall Walton (1978) afirma que esos actos de imaginación por parte de un intérprete complementan el texto del mismo modo que lo hacen los grabados que aparecen en las versiones ilustradas de las novelas, «combinándose con la novela para formar un mundo [ficticio, heterofenomenológico] [mayor]» (pág. 17). Estos añadidos son perfectamente reales, pero no son otra cosa que más «texto», y no hecho de figmento, sino de juicio. No hay nada más que decir sobre la fenomenología.

«¡Pero parece haberla!»

¡Exacto Parece haber fenomenología. Éste es el hecho que, con entusiasmo, admite el heterofenomenólogo. Pero de este hecho innegable y universalmente admitido *no* se deduce que *realmente* haya fenomenología. Ésta es la clave.

«¿Está usted negando que la conciencia sea un *pleno?*»

Efectivamente. Eso es parte de lo que estoy negando. La conciencia es dispersa y está llena de vacíos, y no contiene ni la mitad de lo que las personas creen que contiene.

«Pero, pero...»

Pero la conciencia parece ser un pleno, ¿no?

«¡Sí!»

Estoy de acuerdo; parece ser un pleno; incluso, como dice Edelman, la continuidad parece ser uno de los «hechos más sorprendentes» de la conciencia, pero...

> «Ya sé, ya sé: del hecho de que parezca ser un pleno no se deduce que *sea* un pleno.»

Veo que ya lo ha captado.

> «Pero tengo otro problema con esta habitación de los espejos que usted llama teoría. Usted afirma que es sólo *como* si hubiera un Significador Central, *como* si hubiera un único autor, *como* si hubiera un lugar al que todo va a parar, y yo no entiendo qué pretende usted decir con tantos *como si*.»

Quizás otro experimento mental le sirva para hacer todo esto un poco más digerible. Imagine que visitamos otro planeta y que descubrimos que sus científicos tienen una teoría encantadora: todo objeto físico posee un alma en su interior, y cada alma ama a las demás almas. Así las cosas, los objetos tienden a moverse los unos hacia los otros, empujados por el amor que sus almas internas sienten las unas por las otras. Podemos suponer, además, que nuestros científicos han desarrollado unos sistemas bastante precisos de localización de almas, de modo que, habiendo determinado la localización exacta en el espacio físico del alma de un objeto, pueden responder a preguntas sobre su estabilidad («se va a caer porque su alma está muy alta»), sobre vibraciones («si colocamos un contrapeso a un lado de este engranaje, que tiene un alma muy grande, se reducirán un poco las oscilaciones»), y sobre muchas otras cuestiones técnicas.

Evidentemente, podríamos decirles que habían dado con el concepto de centro de gravedad (o, más precisamente, de centro de masa) y que quizá lo están tratando de manera demasiado ceremoniosa. Les diríamos que pueden seguir hablando y pensando de la manera que lo hacían antes; sólo tienen que suprimir un poco de su innecesario bagaje metafísico. Existe una interpretación más simple y austera (y mucho más satisfactoria) de los hechos que ellos interpretaban a través de una física del alma. Entonces nos preguntarían: ¿existen las almas? Bueno, sí, les contestaríamos, sólo que son *abstractas*, abstracciones matemáticas y no pepitas hechas de alguna sustancia misteriosa. Se trata de ficciones exquisitamente útiles. Es *como* si cada objeto atrayera a cualquier otro, concentrando todo su poder de atracción gravitatoria en un único punto y, además, es mucho más fácil calcular el comportamiento de los objetos utilizando esta ficción de lo que lo sería si nos ocupáramos de todos los detalles, considerando que cada punto atrae a cada punto.

«Me siento *como* si me acabaran de robar la cartera.»

No diga que no le avisé. Usted no puede esperar que la conciencia sea exactamente lo que usted *quiere* que sea. Además, ¿a qué está renunciando?

«Nada más que a mi alma.»

Pero en ningún sentido coherente y defendible del término. Todo aquello a lo que usted está renunciando no es más que una pepita con un carácter muy especial que, por otra parte, tampoco podría ser especial. ¿Por qué debería usted tener una mejor opinión de sí mismo si resultara ser una perla mental dentro de la ostra cerebral? ¿Qué tendría de tan especial el hecho de ser una perla mental?

«Una perla mental, a diferencia de un cerebro, podría ser inmortal.»

La idea de que el yo —o el alma— no es realmente más que una abstracción se le antoja a mucha gente como una idea negativa, como una negación en lugar de algo positivo. Pero de hecho merece mucho la pena adoptarla, incluso —si es que eso le preocupa— porque proporciona una concepción mucho más robusta de la inmortalidad potencial de lo que podemos encontrar en las ideas tradicionales que cualquiera pueda tener sobre el alma, pero esto tendrá que esperar hasta el capítulo 13. Primero, tendremos que ocuparnos de una vez por todas de los qualia, que *aún* tienen un fuerte arraigo en nuestra imaginación.

Capítulo 12

LOS QUALIA DESCALIFICADOS

1. Una nueva cuerda para la cometa

> Si lo lanzamos en un agujero causal, un quale simplemente pasará a través de él.
>
> IVAN FOX, 1989, pág. 82.

Cuando la cuerda de su cometa se enreda, en principio es posible desenredarla, especialmente si usted es persona paciente y analítica. Sin embargo, hay momentos en los que se dejan de lado los principios y triunfa el espíritu práctico. Vaya a comprar una cuerda nueva para la cometa. Le saldrá más barato que dedicar tanto tiempo a recuperar la cuerda vieja, y así pronto podrá volver a ver volar su cometa. Así es como se debe proceder, en mi opinión, con el asunto filosófico de los qualia, un tormentoso lío de experimentos mentales, jergas, juegos de palabras, alusiones a presuntas refutaciones, resultados «recibidos» que deberían ser devueltos al remitente y toda una serie de vías muertas y circunloquios cada vez más barrocos y extravagantes. Algunos nudos es mejor cortarlos, de modo que no me dedicaré a hacer un repaso detallado de toda la bibliografía, aunque es posible encontrar en ella momentos de una profundidad y un ingenio que me han sido de gran utilidad (Shoemaker, 1975, 1981, 1988; White, 1986; Kitcher, 1979; Harman, 1990; Fox, 1989). En el pasado intenté desenredar el lío (Dennett, 1988a), pero ahora pienso que lo mejor es intentar empezar otra vez desde el principio.

No es difícil comprender por qué los filósofos se han hecho un enorme lío con los qualia. Empezaron por donde empezaría todo aquel con un mínimo de sentido: con las intuiciones más fuertes y claras sobre sus propias mentes. Todas estas intuiciones, ¡ay!, constituyen un círculo cerrado de doctrinas que se sostienen entre sí, aprisionando las mentes de los filósofos en el Teatro Cartesiano. Y a pesar de que los filósofos han descubierto las paradojas inherentes a este círculo cerrado de ideas —éste es precisamente el motivo por el cual existe toda esa bibliografía sobre los qualia—, nunca han dispuesto de una *visión alternativa completa* a la que adherirse, de modo que han persistido en fiarse de sus fuertes intuiciones, lo cual siempre los devuelve a esa prisión paradójica. Por eso la bibliografía sobre los qualia se hace cada vez más oscura, sin llegar nunca a un acuerdo. Pero ahora ya

disponemos de una visión alternativa, el modelo de las Versiones Múltiples. Utilizándolo podemos ofrecer una explicación positiva bastante diferente de todas estas cuestiones. Más adelante, en las secciones 4 y 5, podremos detenernos para compararla con las visiones a las que se supone que tiene que sustituir.

En un excelente libro introductorio sobre el cerebro encontramos el siguiente pasaje:

> El «color» en sí mismo no existe en el mundo; sólo existe en los ojos y en el cerebro del observador. Los objetos reflejan la luz a diferentes longitudes de onda, pero estas ondas de luz no tienen color (Ornstein y Thompson, 1984, pág. 55).

Esta afirmación constituye un duro golpe para la visión tradicional, pero nótese que si la tomamos en su sentido estricto y literal, no puede querer decir lo que los autores quieren decir, y no puede ser cierta. El color, dicen los autores, no existe «en el mundo», sino solamente en «los ojos y el cerebro» del observador. Pero los ojos y el cerebro del observador están en el mundo, al igual que lo están otras partes del mundo físico como, por ejemplo, los objetos que ve el observador. Y al igual que esos objetos, los ojos y el cerebro tienen color. Los ojos pueden ser azules, castaños o verdes, e, incluso, el cerebro no está hecho *solamente* de materia gris (y blanca): además de la *substancia nigra* (la sustancia negra), está el *locus ceruleus* (el lugar azul). No obstante, es evidente que los colores que están «en los ojos y en el cerebro del observador» en *este* sentido no son aquello de lo que los autores están hablando. ¿Qué es lo que nos hace pensar que el color existe en cualquier otro sentido?

La ciencia moderna —ésta es la versión más o menos estándar de la historia— ha suprimido los colores del mundo físico, sustituyéndolos por radiaciones electromagnéticas incoloras de longitudes de onda diferentes, que rebotan sobre las superficies que de forma variable reflejan y absorben esas radiaciones. Podría parecer que el color está *ahí afuera*, pero no es así. Está *aquí dentro*, en «los ojos y el cerebro del observador». (Si los autores del pasaje no fuesen unos materialistas tan buenos, probablemente habrían dicho que el color está en la *mente* del observador, lo que los habría salvado de la lectura absurda que acabamos de hacer de sus palabras, pero les habría causado muchos más problemas.) Pero, entonces, si no hay ningún *figmento* interno que pueda colorearse en algún sentido especial, subjetivo, interno a la mente y fenoménico del término, ¡los colores desaparecen por completo! *Algo* tiene que ser esos colores que conocemos y amamos, esos colores que mezclamos y combinamos. ¿Dónde, ay, pueden estar?

Éste es el viejo enigma filosófico con el que tenemos que enfrentarnos. En el siglo XVII, el filósofo John Locke (y antes que él, el científico Robert Boyle) bautizó las propiedades como los colores, los olores, los sabores y los sonidos con el nombre de *cualidades secundarias*, que distinguía de las cua-

lidades primarias como el tamaño, la forma, el movimiento, el número y la solidez. Las cualidades secundarias no eran cosas en la mente, sino *poderes de las cosas* en el mundo (gracias a sus cualidades primarias particulares) para producir y provocar ciertas cosas en las mentes de los observadores. (¿Y qué ocurre cuando no hay ningún observador? Éste es el eterno y conocido enigma del árbol que cae en el bosque. ¿Hace algún ruido? La respuesta queda como ejercicio para el lector.) La definición de las cualidades secundarias que dio Locke se ha convertido en la interpretación más común de la ciencia que hace el profano, y tiene sus virtudes, pero también se cobra sus prendas: las cosas producidas en la mente. La cualidad secundaria de *rojo*, por ejemplo, era para Locke la propiedad disposicional o poder que poseen ciertas superficies de los objetos físicos, gracias a los rasgos microscópicos de su textura, para producir en nosotros la *idea de rojo* cada vez que la luz se reflejaba desde esas superficies hacia nuestros ojos. Puede que el poder del objeto exterior esté bastante claro, así lo parece por lo menos, pero, ¿qué cosa puede ser una idea de rojo? ¿Podemos acaso decir que, como un bonito traje azul, está *coloreada*, en cierto sentido? ¿O, por el contrario, debemos decir que, como una animada discusión sobre el color morado, es *sobre* un color, sin que por ello podamos decir que ella misma está coloreada? Ello nos abre algunas posibilidades, pero, ¿cómo puede ser una idea *sobre* un color (el color rojo, por ejemplo), si no hay nada en ninguna parte que *sea* rojo?

De todos modos, ¿qué es el rojo? ¿Qué son los colores? Los colores siempre han sido el ejemplo favorito de los filósofos y, por el momento, no voy a apartarme de la tradición. En el análisis filosófico de Wilfrid Sellars (1963, 1981b) quien distinguió las propiedades disposicionales de los objetos (las cualidades secundarias de Locke) de lo que él denominó propiedades ocurrentes, encontramos un excelente ejemplo del problema principal con que se encuentra la tradición. Un cubito de hielo de color rosa guardado en la nevera con la luz apagada posee la cualidad secundaria de rosa, pero no existe ninguna instancia de la propiedad de *rosa ocurrente* hasta que el observador abre la nevera y mira en su interior. ¿Es el rosa ocurrente una propiedad de algo en el cerebro o de algo «en el mundo exterior»? En cualquier caso, insiste Sellars, el rosa ocurrente es una propiedad «homogénea» de algo real. Parte de lo que Sellars quería negar con su insistencia en la homogeneidad sería la hipótesis de que el rosa ocurrente es algo parecido a la *actividad neuronal de intensidad 97 en la región 75* del cerebro. También quería negar que el mundo subjetivo de la fenomenología del color se agote con algo tan incoloro como los *juicios* de que una cosa u otra es, o parece ser, de color rosa. Por ejemplo, el acto de recordar en nuestro ojo de la mente el color de un plátano maduro y de juzgar que ese color es el amarillo no tiene en sí mismo la capacidad de provocar la existencia del amarillo ocurrente (Sellars, 1981; Dennett, 1981b). Eso no sería más que juzgar que algo era amarillo, un fenómeno que en sí mismo tiene tan poco amarillo ocurrente como el que tendría un poema sobre los plátanos.

Sellars llegó al extremo de afirmar que todas las ciencias físicas deberían sufrir un cambio revolucionario a fin de acomodar conceptos como el de rosa ocurrente. Muy pocos filósofos le siguieron en esta visión tan radical, aunque, recientemente, el filósofo Michael Lockwood (1989) ha resucitado una versión de la misma. Otros filósofos, como por ejemplo Thomas Nagel, han considerado que incluso después de tal revolución la ciencia seguiría siendo incapaz de tratar con dichas propiedades:

> Los rasgos subjetivos de los procesos mentales conscientes —al contrario que sus causas y efectos físicos— no pueden ser recogidos por la forma de pensamiento purificado apropiada para tratar con el mundo físico que subyace a las apariencias (1986, pág. 15).

Los filósofos han adoptado varios nombres para las cosas en el observador (o propiedades del observador) que se supone que proporcionan cobijo a los colores y el resto de propiedades que han sido expulsadas del mundo «externo» por las victorias de la física: «sensaciones puras», «sensa», «cualidades fenoménicas», «propiedades intrínsecas de las experiencias conscientes», «el contenido cualitativo de los estados mentales» y, por supuesto, «qualia», el término que yo utilizaré aquí. Existen diferencias bastante sutiles entre las definiciones de cada uno de estos términos, pero no me voy a andar con muchos miramientos con ellas. En el capítulo anterior pareció que yo negaba la existencia de *cualquiera* de estas propiedades y, sin que sirva de precedente, esta vez sí es lo que parece: *niego* que cualesquiera de estas propiedades exista. Pero (y aquí vuelve nuestro tema recurrente) estoy completamente de acuerdo en que parece haber qualia.

Parece haber qualia porque realmente parece como si la ciencia nos hubiera demostrado que los colores no pueden estar ahí afuera y que, por tanto, tienen que estar aquí dentro. Además, parece que lo que está aquí dentro no pueden ser *solamente* los juicios que emitimos cuando los cosas nos parecen coloreadas. Sin embargo, este razonamiento es confuso. Lo que la ciencia ha demostrado es simplemente que las propiedades de reflexión de la luz de los objetos hacen que las criaturas entren en diversos estadios discriminativos, esparcidos por sus cerebros y subyacentes a toda una serie de disposiciones innatas y hábitos aprendidos de muy variada complejidad. ¿Y cuáles son sus propiedades? Aquí podemos jugar la carta de Locke por segunda vez: estos estados discriminativos de los cerebros de los observadores poseen varias propiedades «primarias» (las propiedades mecánicas debidas a sus conexiones, los estados de excitación de sus elementos, etc.) y, en virtud de estas propiedades primarias, poseen diversas propiedades secundarias y meramente disposicionales. En las criaturas humanas con lenguaje, por ejemplo, estos estados discriminativos a menudo ponen a estas criaturas en disposición de expresar eventualmente juicios verbales en alusión al «color» de las cosas. Cuando alguien dice «sé que el anillo no es realmente rosa, pero la verdad es que parece rosa», la primera cláusula expresa

un juicio sobre algo en el mundo, y la segunda cláusula expresa un juicio de segundo orden sobre un estado discriminativo sobre algo en el mundo. La semántica de dichos enunciados deja claro cuál es la supuesta naturaleza de los colores: las propiedades de reflexión de las superficies de los objetos o de los volúmenes transparentes (el cubito de hielo rosa o un rayo de luz fluorescente). Y de hecho eso es efectivamente lo que son, aunque determinar exactamente *qué* propiedades de reflexión tienen no es tarea fácil (por motivos que estudiaremos en el próximo apartado).

¿Poseen *también* nuestros estados internos discriminativos propiedades «intrínsecas» especiales, esas propiedades subjetivas, privadas e inefables que constituyen *la manera en que a nosotros nos parece que vemos las cosas* (que oímos, que olemos, etc.)? Estas propiedades adicionales serían los qualia, pero antes de examinar los argumentos que han desarrollado los filósofos en sus intentos de *probar* que existen dichas propiedades adicionales, intentaremos eliminar los motivos para creer en estas propiedades, buscando explicaciones alternativas para los fenómenos que parecen necesitar de ellas. Así, los errores que sistemáticamente se cometen en esos intentos de probar la existencia de los qualia serán más fácilmente perceptibles.

De acuerdo con esta visión alternativa, los colores *son* propiedades que, después de todo, están «ahí afuera». En lugar de las «ideas de rojo» de Locke, tenemos (en los seres humanos normales) estados discriminativos cuyo contenido es: *rojo*. Un ejemplo nos servirá de ayuda para poner en claro qué son, y, aún más importante, qué no son, exactamente estos estados discriminativos. Podemos comparar el color de las cosas del mundo colocándolas las unas al lado de las otras y observándolas, a fin de comprobar a qué conclusión llegamos, pero también comparar colores recordándolos o imaginándolos «en nuestras mentes». ¿Es el rojo de las franjas de la bandera americana igual, más oscuro, más claro, más brillante o más o menos anaranjado que el rojo del traje de Papá Noel (o que el rojo de los buzones en Gran Bretaña, o que el rojo de la estrella roja soviética)? (Si ninguna de estas referencias está disponible en su memoria, inténtelo con otros pares como el azul tarjeta Visa y el azul cielo, el verde mesa de billar y el verde manzana de la Golden, o el amarillo limón y el amarillo canario.) Somos capaces de hacer estas comparaciones «en nuestros ojos de la mente», y cuando lo hacemos, logramos que algo ocurra en nosotros que recupera información de la memoria y nos permite comparar, en la experiencia consciente, los colores de los objetos tal como los recordamos (o, en todo caso, tal como nos parece que los recordamos). Algunos somos mejores que otros en este tipo de ejercicio, sin duda, y muchos de nosotros no nos sentimos muy seguros ante los juicios que emitimos bajo tales circunstancias. Por este motivo llevamos muestras de color o de tela cuando vamos a la tienda, porque así podemos poner las muestras de los colores que queremos comparar en el mundo exterior una junto a otra.

¿Qué ocurre, de acuerdo con mi propuesta, cuando hacemos estas comparaciones «en nuestros ojos de la mente»? Pues algo muy parecido a lo que

ocurriría en una máquina —un robot— que también fuera capaz de efectuar estas comparaciones. Recuérdese el Vorsetzer del CADBLIND Mark I del capítulo 10 (aquel que tenía una cámara que se podía enfocar hacia la pantalla del sistema de CAD). Supóngase que le presentamos una imagen en color de Papá Noel y le preguntamos si el rojo de esta imagen es más oscuro que el rojo de la bandera americana (un dato que ya tiene almacenado en memoria). Esto es lo que haría: recuperaría de la memoria su representación de la Old Glory y localizaría las franjas «rojas» (llevan la etiqueta «rojo #163» en su diagrama). Después comparará este rojo con el rojo del traje de Papá Noel en la imagen que tiene enfrente de la cámara, que el sistema de gráficos en color traducirá como rojo #172. Finalmente, comparará ambos rojos *restando 163 a 172 y obteniendo 9 como resultado*, lo cual podría tener la interpretación, pongamos por caso, de que el rojo de Papá Noel le parece (a *él*) algo más oscuro y más rico que el rojo de la bandera americana.

Esta historia es una simplificación deliberada, que me es útil para poner de manifiesto la afirmación que deseo hacer: es obvio que el CADBLIND Mark I no utiliza el figmento para reproducir su memoria (o su percepción en el momento), *pero nosotros tampoco.* El CADBLIND Mark I probablemente no sabe cómo compara los colores de algo que ve con los de algo que recuerda y *nosotros tampoco.* El CADBLIND Mark I posee —admito esto— un espacio cromático bastante pobre y simple con muy pocas de las asociaciones y tendencias que integran el espacio cromático personal de un ser humano, pero aparte de esta notable diferencia en complejidad disposicional, no hay ninguna otra diferencia importante. Incluso me atrevería a exponerlo así: por lo que se refiere a esta tarea, no hay ninguna diferencia *cualitativa* entre la actuación del CADBLIND y la nuestra. Los estados discriminativos del CADBLIND Mark I tienen contenido del mismo modo, y por los mismos motivos, que los estados cerebrales discriminativos que he colocado en el lugar de las ideas de Locke. El CADBLIND Mark I *no cabe duda* de que carece de qualia (cuando menos, aquí es donde espero que salten los amantes de los qualia), de modo que se sigue de mi comparación que estoy afirmando que nosotros tampoco tenemos qualia. El *tipo* de diferencia que las personas creen que hay entre cualquier máquina y cualquier experimentador humano (recuérdese la máquina catadora de vinos que imaginamos en el capítulo 3) es algo que niego tajantemente: no existe ese tipo de diferencia. Sólo parece haberla.

2. ¿Por qué existen los colores?

Cuando, en el capítulo 11, Otto opinaba que parecía haber un anillo rosado y brillante, ¿cuál era el contenido de este juicio? Si, como he venido insistiendo, ese juicio no era sobre un quale, una propiedad de un aparente anillo «fenoménico» (hecho de figmento), entonces ¿sobre qué era? ¿Qué propiedad se sintió tentado a atribuir (falsamente) a algo en el mundo?

Muchos han observado que, curiosamente, resulta difícil decir exactamente qué tipo de propiedades de las cosas en el mundo podrían ser los colores. La idea más simple y atractiva —la que todavía hallamos en muchas discusiones a nivel elemental— es que cada color puede asociarse a una única longitud de onda de la luz y, por tanto, que la propiedad de ser rojo consiste simplemente en la propiedad de reflejar la luz en la longitud de onda del rojo y de absorber las demás longitudes de onda. Sin embargo, hace ya algún tiempo que se sabe que esto no es correcto. Superficies con propiedades de reflexión fundamentalmente diferentes pueden verse del mismo color, mientras que una misma superficie, bajo condiciones diferentes de luminosidad, puede verse de colores diferentes. Las longitudes de onda de la luz que penetran en nuestros ojos sólo están indirectamente relacionadas con los colores que vemos en los objetos. (Véanse Gouras, 1984; Hilbert, 1987; y Hardin, 1988, donde se estudian diversos aspectos de este asunto con diferente nivel de detalle.) Para aquellos que en algún momento tuvieron la esperanza de hallar una manera simple y elegante de recuperar los beneficios del pagaré que Locke extendió sobre los poderes disposicionales de las superficies, la situación no podría ser peor. Algunos (por ejemplo Hilbert, 1987) han decidido anclar el color de forma objetiva, decidiendo que se trata de una propiedad relativamente simple de los objetos externos, como por ejemplo la propiedad de la «reflectancia espectral de las superficies»; al optar por esta vía, no tienen más remedio que concluir que la visión en color normal a veces nos provoca ilusiones, ya que las constancias que percibimos no siempre coinciden con las constancias de la reflectancia espectral de las superficies medidas con instrumentos científicos. Otros han llegado a la conclusión de que las propiedades de color deben ser consideradas desde un punto de vista subjetivo, como propiedades definibles únicamente en términos de sistemas de estados cerebrales de los observadores, ignorando así la confusa variación en el mundo que da lugar a dichos estados: «Los objetos coloreados son ilusiones, pero no ilusiones infundadas. Normalmente nos hallamos en estados perceptivos cromáticos, los cuales son estados neuronales» (Hardin, 1988, pág. 111; véase Thompson, Palacios y Varela, en prensa, para un examen crítico de estas propuestas y para nuevos argumentos en favor de la opción que presentaremos aquí.)

Lo que ya queda fuera de toda duda es que exista una propiedad simple y no disyuntiva de las superficies tal que sólo aquellas superficies que la poseen, y sólo aquéllas, sean rojas (en el sentido de las cualidades secundarias de Locke). En principio, éste es un hecho turbador, casi deprimente, ya que parece indicar que nuestra aprehensión perceptiva del mundo es mucho peor de lo que habíamos pensado, como si habitáramos un mundo de sueños o como si fuéramos víctimas de un engaño colectivo. Nuestra visión en color no nos proporciona un acceso a las propiedades simples de los objetos, aunque parezca lo contrario. ¿Por qué?

¿Mala suerte? ¿Un diseño defectuoso? Por supuesto que no. Existe una perspectiva diferente, y mucho más instructiva, que podemos adoptar en

cuanto al color, que me fue mostrada por primera vez por la filósofa de las neurociencias Kathleen Akins (1989, 1990).[1] A veces, existe una razón por la cual han aparecido nuevas propiedades. Un ejemplo particularmente útil es el que nos proporciona el famoso caso de Julius y Ethel Rosenberg, que fueron condenados y ejecutados en 1953 por espiar el proyecto estadounidense de la bomba atómica en favor de la Unión Soviética. Durante el proceso se desveló que habían improvisado un inteligente sistema de identificación: rompían en dos trozos una caja de cartón de gelatina Jell-O, y cada trozo se remitía a los dos individuos, que debían ir con mucho cuidado en el momento de identificarse. Cada trozo se convertía así en un «detector» único y casi infalible del compañero: en futuros encuentros, cada parte debía mostrar su trozo y si ambas partes encajaban, todo iba bien. ¿Por qué funciona este sistema? Porque al romper el cartón se produce un perfil de tal complejidad informacional que sería virtualmente imposible reproducir de forma deliberada. (Nótese que cortar la caja de gelatina con una cuchilla no serviría para los propósitos que hemos descrito.) El borde irregular de un trozo de cartón se convierte en un dispositivo de reconocimiento de patrones *prácticamente* único de su pareja; es un aparato o un traductor para detectar la propiedad formal *M*, donde *M* se instancia sólo en su pareja.

En otras palabras, la propiedad formal *M* y el detector de la propiedad *M* que la detecta están hechos el uno para la otra. Si el uno o la otra no existiera, tampoco habría ningún motivo para que existiera la otra parte, no habría motivos para que fuera creada. Lo mismo ocurre con los colores y la visión en color: están hechos los unos para la otra. Los códigos de colores son una idea bastante reciente en la «ingeniería de los factores humanos», pero sus virtudes ya han sido ampliamente reconocidas. En los hospitales se trazan líneas de colores por los pasillos, lo cual ayuda a los pacientes a orientarse: «Para llegar a fisioterapia, siga la línea amarilla»; para llegar al banco de sangre, siga la línea roja». Los fabricantes de televisores, ordenadores y otros aparatos electrónicos utilizan un código de colores para los haces de cables a fin de poder seguir el recorrido del cable de un punto a otro. Estas aplicaciones son muy recientes, pero la idea es mucho más antigua; más antigua que la letra escarlata con que se marcaba a las adúlteras, más antigua que los uniformes de colores que ayudaban a distinguir al amigo del enemigo en el fragor de la batalla, más antigua, de hecho, que la misma especie humana.

Tenemos la tendencia a pensar en los códigos de colores como una inteligente introducción de esquemas «convencionales» diseñados para sacarle partido a la visión en color «natural»; pero esta idea no tiene en cuenta el hecho de que la visión en color «natural» coevolucionó *desde el principio* con los colores, cuya *raison d'être* no era otra que la de establecer códigos de colores (Humphrey, 1976). En la naturaleza, algunas cosas «necesitaban

1. Podemos hallar variaciones sobre este mismo tema en Humphrey (1976, 1983a) y Thompson, Palacios y Varela (1992).

ser vistas», mientras que otras necesitaban verlas, de modo que evolucionó un sistema que tendía a minimizar el trabajo de las segundas, potenciando la capacidad de resaltar de las primeras. Considérense los insectos. Su visión del color coevolucionó con los colores de las plantas que polinizaban, un buen truco de diseño que benefició a ambos. Sin el código de colores de las flores, la visión en color de los insectos no habría evolucionado, y viceversa. Así pues, el principio del código de colores es la base de la visión en color de los insectos, y no una invención reciente de alguna especie inteligente de mamíferos. Podríamos contar historias similares sobre la evolución y la visión en color en otras especies. Mientras que es posible que algún tipo de visión en color haya evolucionado con el propósito de discriminar visualmente ciertos fenómenos inorgánicos, no está claro que esto se haya producido con ninguna de las especies de este planeta. (Evan Thompson me ha hecho notar que las abejas podrían utilizar su tipo especial de visión en color para la navegación, a fin de discriminar la luz solar polarizada en los días nublados, pero, ¿no sería éste un uso secundario de la visión en color que coevolucionó originalmente con los colores de las flores?)

Los diferentes sistemas de visión en color han evolucionado de forma independiente, en ocasiones con espacios cromáticos radicalmente distintos. (Para un breve examen de estas cuestiones y referencias bibliográficas adicionales, véase Thompson, Palacios y Varela, 1992). No todas las criaturas vivientes poseen algún tipo de visión en color. Los pájaros, los peces, los reptiles y los insectos poseen una visión en color muy parecida a nuestro sistema «tricromático» (rojo-verde-azul); los perros y los gatos no. Entre los mamíferos, sólo los primates poseen visión en color, y con diferencias sorprendentes entre los diferentes sistemas. ¿Qué especies poseen visión en color y por qué? Esta historia resulta ser fascinante y compleja y, en gran medida, llena de especulaciones.

¿Por qué las manzanas son rojas cuando han madurado? Es natural suponer que podemos dar una respuesta únicamente en términos de cambios químicos que se producen cuando el azúcar y otras sustancias alcanzan unos determinados grados de concentración en la fruta durante el proceso de maduración, produciendo así reacciones diversas. Pero ello ignora el hecho de que no habría manzanas si no hubiera comedores de manzanas y esparcidores de semillas que pudieran verlas, de modo que el hecho de que las manzanas sean fácilmente visibles para al menos algunas variedades de comedores de manzanas constituye una condición para su propia existencia, y no una mera «casualidad» (¡desde el punto de vista de las manzanas!). El hecho de que las manzanas posean la reflectancia espectral de las superficies que poseen es tanto una función de los fotopigmentos que había disponibles para equipar las células cónicas en los ojos de los fructívoros como lo es de los efectos de las interacciones entre el azúcar y los otros componentes de la química de la fruta. Los frutos que no participan del código de colores compiten muy mal en los estantes del supermercado de la naturaleza, aunque la publicidad engañosa también será penalizada; los frutos que es-

tán maduros (llenos de nutrientes) *y que lo publicitan* se venderán mejor, pero la publicidad debe acomodarse a las capacidades visuales y a las inclinaciones de los potenciales consumidores.

Al principio los colores se hicieron para que pudieran verlos aquellos que estaban hechos para verlos. Pero esta situación fue evolucionando gradualmente, por casualidad, aprovechando hábilmente los materiales que estaban a mano, estallando ocasionalmente en una profusión de elaboraciones de un nuevo truco, y siempre tolerando altos grados de variación *y* de constancia inútil (meramente coincidente). Dichas constancias coincidentes a menudo afectaban a rasgos «más fundamentales» del mundo físico. Una vez hubo criaturas capaces de distinguir las bayas rojas de las bayas verdes; éstas también pudieron distinguir los rubíes rojos de las esmeraldas verdes, pero eso no fue más que una bonificación fruto de la coincidencia. El hecho de que exista una diferencia *de color* entre los rubíes y las esmeraldas puede, pues, considerarse como un fenómeno cromático *derivado*. ¿Por qué es azul el cielo? Porque las manzanas son rojas y las uvas moradas, pero no al revés.

Es un error pensar que primero hubo colores —rocas coloreadas, agua coloreada, cielo coloreado, orín rojo-anaranjado y cobalto azul brillante— y que después la Madre Naturaleza apareció y supo sacar partido de *esas* propiedades para fijar un código de colores para las cosas. Por el contrario, primero había las diferentes propiedades reflectantes de las superficies, propiedades reactivas de los fotopigmentos, etc., y después la Madre Naturaleza desarrolló, a partir de estas materias primas, unos eficientes sistemas de codificación en «color»/de visión en «color» ajustados el uno con el otro, y entre las propiedades que surgieron de este proceso de diseño se encuentra lo que los seres humanos normales llamamos colores. Si resultara que el azul del cobalto y el azul de un ala de mariposa fueran iguales (para la visión de un ser humano normal), ello no sería más que una coincidencia, un efecto secundario desdeñable, fruto de los procesos que condujeron al nacimiento de la visión en color, y así (como el propio Locke habría reconocido), dotaron a un conjunto curiosamente amañado de propiedades primarias complejas con la propiedad secundaria compartida de producir un efecto común en un conjunto de observadores normales.

«Y sin embargo», apuntará usted, «¡antes de que hubiera animales con visión en color, ya había gloriosas puestas de sol rojas, y brillantes esmeraldas verdes!» Sí, claro, usted puede decir eso, pero entonces esas mismas puestas de sol también serían chillonas, multicolores y desagradables, reproducidas en colores que no podemos ver y para los cuales ni siquiera tenemos un nombre. Es decir, usted no podrá más que admitir esto, si hubiera o *pudiera haber* criaturas en algún planeta cuyo aparato sensorial se viera afectado de cualquiera de estos modos por dichas puestas de sol. Y por lo que sabemos, existen especies que perciben naturalmente que hay dos (o diecisiete) colores diferentes en un puñado de esmeraldas que *nosotros* vemos uniformemente de color verde.

Muchos seres humanos son ciegos a los colores rojo y verde. Suponga-

mos que todos lo fuéramos, en cuyo caso estaríamos de acuerdo en que tanto los rubíes como las esmeraldas son «rojerdes»; después de todo, los demás observadores normales las ven igual que muchas otras cosas rojerdes: los coches de bomberos, los céspedes bien regados, las manzanas maduras y las manzanas no maduras (Dennett, 1969). Si aparecieran unos individuos como nosotros, insistiendo en que las esmeraldas y los rubíes son de colores diferentes, no habría manera de decidir cuál de los dos sistemas de visión en color es más «fiel».

El filósofo Jonathan Bennett (1965) llama nuestra atención sobre un caso que plantea una situación parecida, quizá más persuasiva, dentro de una modalidad sensorial diferente. La sustancia *fenol-thio-urea*, nos cuenta Bennett, sabe amarga para un cuarto de la población humana y es totalmente insípida para el resto. El sabor que usted percibe es algo que está determinado genéticamente. ¿La *fenol-thio-urea* es amarga o insípida? Por «eugenesia» (reproducción controlada) o por ingeniería genética, podríamos llegar a eliminar el genotipo para percibir amargura en la mencionada sustancia. Si lo consiguiéramos, la *fenol-thio-urea* sería entonces *paradigmáticamente* insípida, como el agua destilada: insípida para todos los seres humanos. Ahora bien, antes de que hubiera seres humanos, ¿era la *fenol-thio-urea a la vez* amarga e insípida? Químicamente era lo mismo que es ahora.

Los hechos sobre las cualidades secundarias están ineludiblemente ligados a la referencia a una determinada clase de observadores, pero existen maneras más o menos fuertes de tratar ese vínculo. Podríamos decir que las cualidades secundarias son *encantadoras* en vez de *sospechosas*. Una joven podría ser encantadora sin que ningún observador susceptible de encontrarla encantadora la hubiera observado antes, pero no podría —por lógica— ser sospechosa hasta que alguien empezara a sospechar de ella. Podría decirse que existen ejemplos particulares de cualidades encantadoras (como la cualidad misma de ser encantador) en tanto que disposiciones lockeanas, con anterioridad al momento (si es que éste llega) en que ejercen su poder sobre algún observador, produciendo el efecto definitorio en él. Así pues, una mujer que nadie hubiera visto antes nunca (que hubiera crecido sola en una isla desierta, supongo) podría ser realmente encantadora, poseer el poder disposicional de afectar de una determinada manera a unos observadores normales de una determinada clase, a pesar de no haber tenido nunca la oportunidad de hacerlo. Pero las cualidades encantadoras no pueden definirse independientemente de las inclinaciones, susceptibilidades o disposiciones de una clase de observadores, de modo que no tiene sentido hablar de la existencia de cualidades encantadoras de forma totalmente independiente de la existencia de los observadores pertinentes. Esto es, de hecho, demasiado fuerte. Las cualidades encantadoras no se *definirían* —no tendría ningún sentido definirlas, al contrario que otras propiedades amañadas lógicamente posibles— independientemente de dicha clase de observadores. Así que, mientras podría ser lógicamente posible («retrospectivamente» podríamos decir) reunir ejemplos de propiedades cromáticas me-

diante un mero proceso de enumeración, los motivos para distinguir dichas propiedades (por ejemplo, para explicar ciertas regularidades causales dentro de un conjunto de unos objetos curiosamente complicados) dependen de la existencia de la clase de los observadores.

¿Son encantadores los elefantes marinos? Para nosotros no. Es difícil imginar una criatura más fea. Lo que hace que un elefante marino resulte encantador para otro elefante marino no es lo mismo que hace que una mujer resulte encantadora para un hombre, y calificar de encantadora a una mujer nunca vista que ejerciera un poderoso poder de atracción sobre los elefantes marinos sería un abuso, tanto del término como de la mujer. Es sólo por referencia al gusto de los humanos, que es un rasgo contingente e idiosincrásico del mundo, que la propiedad de ser encantador (para un ser humano) puede ser identificada.

Por otra parte, las cualidades sospechosas (tales como la cualidad de ser sospechoso) se entienden de tal modo que se debe presuponer que cualquier ejemplo de dicha propiedad ya ha tenido ocasión de mostrar su efecto definitorio sobre un observador cuando menos. Usted puede ser totalmente digno de sospecha —usted puede ser incluso de referencia claramente culpable—, pero usted no puede ser un sospechoso hasta que alguien sospeche de usted. No estoy afirmando que los colores sean cualidades sospechosas. Nuestra intuición de que toda esmeralda nunca vista, oculta en su lecho mineral, *ya* es verde no tiene por qué negarse. Sí estoy afirmando, en cambio, que los colores son cualidades encantadoras, cuya existencia, ligada como está a una clase de observadores de referencia, no tiene sentido en un mundo en el que no haya lugar para los observadores. Esto es mucho más fácil de aceptar para algunas cualidades secundarias que para otras. El hecho de que los vapores sulfurosos que despedían los volcanes primordiales fueran amarillos parece ser un poco más objetivo que el hecho de que apestaran, pero en tanto en cuanto lo que queremos decir con «amarillo» es lo que *nosotros* queremos decir con «amarillo», ambas afirmaciones son equivalentes. Supongamos ahora que un terremoto primordial causara una fractura en el terreno dejando al descubierto cientos de franjas de estratos químicamente diferentes. ¿Eran *visibles* estas franjas? ¿A quién se lo tenemos que preguntar? Quizás algunas fuesen visibles para nosotros y otras no. Quizás algunas de las franjas invisibles para nosotros fuesen visibles para las palomas tetracrómatas, o para criaturas capaces de ver en la parte infrarroja o ultravioleta del espectro electromagnético. Por el mismo motivo, no tiene sentido preguntar si la diferencia entre las esmeraldas y los rubíes es una diferencia visible, si no se especifica primero el sistema de visión.

La evolución suaviza el «subjetivismo» o «relativismo» implícito en el hecho de que las cualidades secundarias son cualidades encantadoras. Demuestra que la ausencia de rasgos comunes «simples» o «fundamentales» en cosas que son del mismo color no es una señal ineludible de ilusión total, sino

un signo de amplia tolerancia por «detecciones positivas falsas» de las propiedades ecológicas que realmente importan.[2]

Las categorías básicas de nuestros espacios cromáticos (y, evidentemente, de nuestros espacios olfativos, nuestros espacios sonoros y todo lo demás) han sido formadas por presiones selectivas, de modo que por lo general tiene sentido preguntar para qué sirve una determinada discriminación o preferencia. Existen motivos por los cuales rechazamos ciertos olores y aceptamos otros, por los cuales preferimos ciertos colores a otros, por los cuales ciertos sonidos nos molestan más que otros, o nos relajan más. No tienen por qué ser siempre nuestros motivos, sino que pueden ser los motivos de unos antepasados lejanos, que han dejado sus huellas fósiles en las predisposiciones innatas que conforman nuestros espacios cualitativos. Pero, como buenos darwinianos, también deberíamos reconocer la posibilidad —la necesidad, de hecho— de que existan otras predisposiciones no funcionales, distribuidas al azar por la población gracias a la variación genética. A fin de que la presión selectiva favorezca de forma diferenciada a aquellos que muestren una predisposición en contra de *F* una vez que *F* se ha hecho ecológicamente importante, tienen que haberse producido variaciones inútiles (no funcionales todavía) en las «actitudes hacia *F*» sobre las cuales pudiera actuar la selección. Por ejemplo, si en el futuro comer callos *fuera* a lanzar una maldición prerreproductiva, solamente los que estuviesen «naturalmente» (y, *hasta ese momento*, inútilmente) predispuestos en contra de comer callos tendrían una ventaja (quizá muy leve en un principio, pero pronto explosiva, si las condiciones la favoreciesen). Así pues, si usted encuentra que algo (el brécol, por ejemplo) es indescriptible e inefablemente repugnante, de ello no se sigue que exista algún motivo por el cual esto deba ser así. Ni tampoco se sigue que usted tenga algún defecto en relación a sus semejantes, si no coinciden con usted en cuanto a este punto. Podría ser uno de esos detalles innatos dentro de su espacio cualitativo que, por el momento, no tiene ninguna significación funcional. (Y por su bien, mejor que si llega tenerla, que sea porque el brécol se ha convertido de repente en algo pernicioso para nosotros.)

2. Actualmente, los filósofos sienten cierta afición por el concepo de clase *natural*, reintroducido en la filosofía por Quine (1969), quien ahora lamentaría el que éste se haya convertido en un sinónimo del dudoso, pero secretamente popular, concepto de *esencia*. «Las cosas verdes, o cuando menos las esmeraldas verdes, son una clase», observa Quine (pág. 116), manifestando su propio reconocimiento del hecho de que mientras las esmeraldas verdes pueden ser una clase natural, las cosas *verdes* probablemente no lo son. La presente discusión pretende anticiparse a cualquier tentación de caer en los errores del naturalismo de salón: la idea de que cualquier cosa que produce la naturaleza es una clase natural. Los colores no son «clases naturales» precisamente porque son un producto de la evolución biológica, la cual tiene una tolerancia con los límites difusos cuando produce categorías que horrorizarían a cualquier filósofo amante de las definiciones buenas y claras. Si la vida de alguna criatura dependiera de reunir la luna, el queso azul y las bicicletas en una sola categoría, pueden estar seguros de que la Madre Naturaleza encontraría el modo de que esta criatura «viera» todas estas cosas «intuitivamente como la misma clase de cosas».

Estas consideraciones evolutivas nos ayudan en la tarea de explicar por qué las cualidades secundarias resultan ser tan «inefables» y tan difíciles de definir. Igual que la propiedad formal *M* del trozo de caja de cartón de los Rosenberg, las cualidades secundarias se muestran extremadamente resistentes a una definición simple. Forma parte de la esencia del truco de los Rosenberg el que no podamos sustituir nuestro predicado postizo por una descripción de la propiedad que sea más larga y más compleja, pero a la vez más precisa y exhaustiva, ya que, si pudiéramos hacerlo, nosotros (o cualquier otro) podríamos utilizar dicha descripción como receta para producir otro ejemplo de *M* u otro detector de *M*. Nuestros detectores de cualidades secundarias no fueron diseñados específicamente para detectar solamente propiedades difíciles de definir, pero el resultado es prácticamente el mismo que si lo hubieran sido. Como observa Akins (1989), la razón de ser de nuestros sistemas sensoriales *no es* la de detectar propiedades «básicas» o «naturales» del entorno, sino la de servir a nuestros propósitos «narcisistas» de permanecer con vida; la naturaleza no construye motores epistémicos.

La única manera *fácilmente accesible* de decir cuál es la propiedad formal *M* consiste en señalar hacia el detector de *M* y decir que *M* es la propiedad formal detectada por esa cosa de ahí. Naturalmente, cualquiera que intente decir qué propiedad detecta (o no consigue detectar) alguien cuando «aparece de la manera que a él le parece» se encontrará en el mismo apuro. Ahora podemos responder, por tanto, a la pregunta que planteábamos al principio de este apartado: ¿Qué propiedad opina Otto que tiene algo cuando opina que es de color rosa? La propiedad que él llama color rosa. ¿Y qué propiedad es ésa? Es difícil decirlo, pero eso no debe preocuparnos, porque podemos explicar por qué es tan difícil decirlo. Lo mejor que podemos hacer, en la práctica, cuando se nos pregunta qué propiedades de las superficies detectamos con la visión en color, es decir, mostrándonos muy poco informativos, que detectamos las propiedades que detectamos. Si alguien desea una respuesta más informativa sobre estas propiedades, existe una extensa, y difícilmente resumible, bibliografía en biología, neurociencia y psicofísica que puede consultar. Y Otto no puede decir nada más sobre la propiedad que él llama color rosa que «¡es *ésta!*» (lo que vendría a significar que él está señalando hacia «adentro», hacia una propiedad fenoménica privada de su experiencia). Todo lo que consigue con esto es (como máximo) señalar hacia su propio e idiosincrásico estado de discriminación cromático, lo cual equivale a sostener un trozo de caja de cartón y decir que detecta esa propiedad formal. Otto señala, quizá, hacia su propio dispositivo de discriminación, pero no hacia un quale que dicho dispositivo haya destilado, o que el dispositivo lleve consigo, o que el dispositivo haya reproducido, en el momento de hacer su trabajo. Eso no existe.

> «Y, sin embargo [insiste Otto], usted todavía no ha explicado por qué el color rosa se ve *así*.»

¿Cómo?

> «*Así.* Como ese intrínseco carácter rosado particularmente inefable y maravilloso de que ahora estoy disfrutando. *Eso* no es una complicada e indescriptibe propiedad de reflectancia de las superficies de los objetos externos.»

Veo, Otto, que usted utiliza el término *disfrutar.* Y no es el único. Es frecuente que cuando un autor quiere hacer hincapié en el hecho de que el tema ha pasado de la (mera) neuroanatomía a la experiencia, de la (mera) psicofísica a la conciencia, de la (mera) información a los qualia, la palabra «disfrutar» entre en escena.

3. Disfrutando de nuestras experiencias

> Pero Dan, ¡si los qualia son lo que hace que merezca la pena vivir la vida!
> WILFRID SELLARS (ante una botella de excelente Chambertin, Cincinnati, 1971)

> Si cuando bebo un buen vino lo que quiero es información sobre sus propiedades químicas, ¿por qué no me limito a leer la etiqueta?
> SYDNEY SHOEMAKER, Coloquio en Tufts, 1988.

Algunos colores se hicieron para gustar, así como algunos olores y algunos sabores. Otros colores, olores y sabores, en cambio, se hicieron para no gustar. En otras palabras, y más precisamente, no es un accidente el que a nosotros (y a las demás criaturas capaces de detectarlos) nos gusten y nos disgusten los colores, los olores, los sabores y otras cualidades secundarias. Del mismo modo que hemos heredado unos evolucionados detectores de simetrías verticales en nuestros sistemas visuales para alertarnos (a nosotros y a nuestros antepasados) del hecho ecológicamente significativo de que hay otra criatura que nos está mirando, también hemos heredado unos evolucionados detectores de cualidades secundarias que no son unos meros testimonios desinteresados, al contrario, son avisadores y alertadores, sirenas, tanto en el sentido homérico del término como en el sentido del coche de bomberos.

Como vimos en el capítulo 7, sobre la evolución, estos alarmistas nativos, por cooptación, han pasado, después, a formar parte de toda una serie de complejas organizaciones, construidas a partir de millones de asociaciones, y formadas, en el caso de los humanos, por miles de memas. Así, la atracción animal por el sexo y la comida —ven y cógelo— y la aversión animal por el dolor y el miedo —corre por tu vida— se mezclaron en diferentes proporciones para dar lugar a toda una serie de picantes combinaciones. Cuan-

do un organismo descubre que resulta beneficioso prestar atención a una determinada característica del mundo *a pesar de* su aversión congénita a hacerlo, debe construir una coalición que sirva para contrarrestar e impedir, así, que venza la aversión. La tensión semiestable resultante puede convertirse entonces en un gusto adquirido, algo que se buscará en determinadas circunstancias. Cuando un organismo descubre que tiene que ahogar los efectos de un insistente grupo de alertadores, si quiere seguir el curso adecuado, puede que deba cultivar un gusto por todas aquellas secuencias de actividad que pueda encontrar que tiendan a producirle la paz y la calma que desea. Es de este modo que puede llegar a gustarnos esa comida picante que nos quema la boca (Rozin, 1982), la música deliciosamente «discordante», así como el calmado y frío realismo de Andrew Wyeth a la vez que el inquietante y cálido expresionismo de Willem de Kooning. Marshall McLuhan (1967) proclamó que el medio es el mensaje, una media verdad que es más verdad quizás en el caso del sistema nervioso que en cualquier otro foro de comunicación. Lo que queremos cuando tomamos un sorbo de buen vino no es, evidentemente, información sobre su contenido químico; lo que queremos es que *se nos informe* sobre su contenido químico pero de la manera que a nosotros nos gusta. Y nuestras preferencias se basan, *en última instancia*, en las predisposiciones que aún están configuradas en nuestros sistemas nerviosos, aunque su valor ecológico se haya desvanecido hace ya muchísimos años.

Nuestra propia tecnología nos ha venido ocultando este hecho desde hace mucho tiempo. Como señala el psicólogo Nicholas Humphrey,

> Al mirar a mi alrededor, los colores fabricados por el hombre me llaman la atención desde todos los rincones de mi despacho: libros, cojines, una alfombra en el suelo, una taza de café, una caja de grapas —azules, rojos, amarillos y verdes brillantes—. Hay más colores aquí que en una jungla tropical. Y sin embargo, mientras que en la jungla cada color tendría un significado, aquí en mi estudio prácticamente nada lo tiene. La anarquía de los colores ha tomado el poder (1983a, pág. 149).

Considérese, por ejemplo, el curioso hecho de que a los monos no les gusta la luz roja. Si se le deja elegir, un macaco muestra un fuerte preferencia por el extremo azul-verde del espectro, y se pone muy nervioso si tiene que pasar largos períodos en entornos donde domina la luz roja (Humphrey, 1972, 1973, 1983a; Humphrey y Keeble, 1978). ¿Por qué? Humphrey señala que el rojo siempre se ha utilizado como alarma, el color fundamental dentro del código de colores, pero, precisamente por ese motivo, también el más ambiguo: la fruta roja puede ser buena para comer, pero la serpiente o el insecto rojos probablemente están anunciando que son venenosos. El rojo transmite mensajes «mixtos». ¿Pero por qué envía un mensaje de «alerta»? Quizá porque es el contraste más fuerte disponible en un ambiente cuyo fondo es el verde de la vegetación o el azul del mar, o —en el caso de los monos—

porque la luz roja (roja, rojo-anaranjada o anaranjada) es la luz del crepúsculo y del alba, las horas del día en que casi todos los depredadores de los monos salen a cazar.

Las propiedades afectivas o emocionales del rojo no se restringen a los macacos. Todos los primates muestran este tipo de reacciones, incluidos los humanos. Si los trabajadores de su empresa se entretienen demasiado en los lavabos, pinte las paredes de rojo y resolverá el problema, aunque creará otros (véase Humphrey, 1992). Evidentemente, estas reacciones «viscerales» no se restringen a los colores. La mayoría de los primates que han crecido en cautividad y que nunca han visto una serpiente demostrarán inmediatamente que las odian tan pronto como vean una, y es muy probable que el tradicional disgusto que sienten los humanos por las serpientes tenga un origen biológico que explicaría el origen bíblico, y no al revés.[3] Es decir, nuestra herencia genética frustra el desarrollo de aquellos memas que no favorezcan el odio por las serpientes.

He aquí dos explicaciones diferentes de la inquietud que la mayoría de nosotros sentimos (incluso si la «dominamos») cuando vemos una serpiente:

1. Las serpientes evocan en nosotros un determinado quale intrínseco de la asquerosidad de las serpientes cuando las miramos, y nuestra inquietud es una reacción a ese quale.
2. No nos suele apetecer mucho ver serpientes a causa de las predisposiciones innatas configuradas en nuestros sistemas nerviosos. Éstas favorecen la liberación de adrenalina, ponen en marcha los mecanismos de lucha-o-huye y, al activar una serie de vínculos asociativos, traen a colación toda una serie de situaciones de peligro, violencia y daño. La aversión original de los primates, en nosotros se ha visto transformada, revisada y desviada por la acción de cientos de memas que la han explotado, cooptado, conformado. (Existen muchos niveles diferentes en los que podríamos desarrollar una explicación «funcionalista» como ésta. Por ejemplo, nos podríamos permitir la libertad de hablar de forma más casual sobre el poder de la percepción de serpientes para provocar ansia, miedo, anticipaciones de dolor, etc., pero ello sería visto como «hacer trampas», y por eso lo evito.)

El problema con el primer tipo de explicación es que sólo parece ser una explicación. Es inútil pensar que la idea de que una propiedad «intrínseca» (de rosa ocurrente, de asquerosidad de las serpientes, de dolor, de aroma de café) podría *explicar* las reacciones de un sujeto a una determinada circunstancia: es un excelente ejemplo de *virtus dormitiva* (véase página 75). No es tarea fácil, sin embargo, el acusar a una teoría de albergar una vacua *virtus*

3. La primatóloga Sue Savage-Rumbaugh me ha informado de que, al contrario que los chimpancés, los bonobos, o chimpancés enanos, que han crecido en cautividad no muestran un miedo innato a las serpientes.

dormitiva. A veces tiene sentido postular una *virtus dormitiva* temporal, en espera de desarrollos futuros. La concepción es, podríamos decir que por definición, la causa del embarazo. Si no dispusiéramos de ninguna otra manera de identificar la concepción, decir a una mujer que está embarazada porque ha concebido sería una afirmación vacua, no una explicación. Pero una vez que hemos desentrañado la teoría mecánica de la concepción, podemos ver *de qué modo* la concepción es la causa del embarazo, y el carácter informativo de nuestra afirmación se ve restaurado. En la misma línea de razonamiento, podríamos identificar (a grandes rasgos) los qualia, por definición, como las causas proximales de nuestro disfrute y de nuestro sufrimiento, y seguidamente proceder a descargarnos de nuestras obligaciones de informar adoptando el otro estilo de explicación. Curiosamente, sin embargo, los qualófilos (así es como llamo a los que todavía creen en los qualia) no tienen ninguna de las dos; insisten, como Otto, en que los qualia «reducidos» a meros complejos de disposiciones a reaccionar realizados de forma mecánica no son los qualia de que ellos hablan. *Sus* qualia son algo diferente.

> «Considere [dice Otto] cómo se me aparece el anillo rosado ahora, en este preciso momento, independientemente de todas mis disposiciones, asociaciones pasadas y actividades futuras. *Esto*, el *modo* puro y aislado *en que se me presenta* en relación al color en este momento, esto es mi quale rosa.»

Otto acaba de cometer un error. De hecho, éste es el gran error, el origen de todas las paradojas sobre los qualia, como enseguida veremos. No obstante, antes de exponer el sinsentido que comporta seguir por esta vía, quiero mostrar algunos aspectos positivos de la vía que Otto rechaza: la vía «reduccionista» de *identificar* «el modo en que se me presenta» con la suma total de todas las disposiciones a reaccionar idiosincrásicas inherentes a mi sistema nervioso como resultado del hecho de que yo me enfrente a un determinado patrón de estímulos.

Considérese lo que debe de haber sido para un pío y devoto luterano de Leipzig en, pongamos por caso, 1725, el escuchar una de las cantatas corales de J. S. Bach el día de su estreno. (Este ejercicio de imaginar *lo que se siente* es un precalentamiento para el capítulo 14, donde nos ocuparemos de la conciencia en otros animales.) Probablemente, no hay ninguna diferencia biológica entre nosotros y los luteranos alemanes del siglo XVIII; pertenecemos a la misma especie, y apenas ha pasado el tiempo. Pero, a causa de la tremenda influencia de la cultura —la memosfera—, nuestro mundo psicológico es muy diferente del suyo, en sentidos que tendrían un profundo impacto en nuestras respectivas experiencias al escuchar una cantata de Bach por primera vez. Nuestra imaginación musical se ha enriquecido y complicado de muchas maneras (por Mozart, por Charlie Parker, por los Beatles), pero también ha perdido muchas de las poderosas asociaciones con las que

podía contar Bach. Sus cantatas corales estaban construidas en torno a las corales, melodías de himnos tradicionales que les eran muy familiares a los devotos y que, por tanto, provocaban oleadas de asociaciones emotivas y temáticas tan pronto como sus huellas o sus ecos aparecían en la música. La mayoría de nosotros conoce estas corales sólo a través de la música de Bach, de modo que cuando las escuchamos, lo hacemos con unos oídos diferentes. Si queremos imaginarnos lo que sentía un ciudadano de Leipzig al escuchar a Bach, no basta con que escuchemos los mismos tonos tocados con los mismos instrumentos en el mismo orden; también tenemos que prepararnos para responder a esos tonos con las mismas angustias, emociones y oleadas de nostalgia.

No es completamente imposible prepararse de esta manera. Un musicólogo que evitara cuidadosamente todo contacto con la música posterior a 1725 y que estuviera profundamente familiarizado con la música tradicional de aquel período sería una buena primera aproximación. Aún más importante, como demuestran estas observaciones, no es imposible saber exactamente cómo deberíamos prepararnos tanto si nos interesara pasar por todo el proceso como si no. Así que podríamos saber lo que se sentiría «en abstracto» por así decir; de hecho, acabo de explicarlo: los ciudadanos de Leipzig, al escuchar las cantatas corales, traían a la memoria todas las asociaciones que ya alimentaban su reconocimiento de las melodías corales. Es bastante fácil imaginar lo que *esto* les debió haber hecho sentir, aunque con variaciones extraídas de nuestra propia experiencia. Podemos imaginar cómo sería escuchar los arreglos de Bach sobre algunas canciones navideñas tradicionales o sobre «Home on the Range». No podemos llevar a cabo el trabajo en toda su extensión, pero solamente porque no podemos olvidar o abandonar todo lo que sabemos y que los ciudadanos de Leipzig no sabían.

A fin de apreciar hasta qué punto es importante este bagaje adicional nuestro, imaginemos que un musicólogo descubriera una cantata de Bach desconocida, definitivamente atribuible al gran músico, oculta en un baúl y, probablemente, nunca interpretada ni escuchada por el propio compositor. Todo el mundo ardería en deseos de escucharla, para experimentar por primera vez los «qualia» que habrían conocido los ciudadanos de Leipzig, si la hubiesen podido escuchar, pero ello resulta finalmente imposible, ya que el tema principal de la cantata es, por una desafortunada coincidencia, ¡las primeras siete notas de «Rudolph the Red-Nosed Reindeer»! Aquellos que, como yo, tenemos los oídos llenos de esta melodía nunca seremos capaces de escuchar la versión de Bach tal como él hubiera querido o tal como los ciudadanos de Leipzig la hubieran recibido.

Sería difícil encontrar un caso más claro de bloqueo de la imaginación, pero nótese que no tiene nada que ver con las diferencias biológicas o incluso con las propiedades «intrínsecas» e «inefables» de la música de Bach. El único motivo por el cual no podríamos revivir de forma imaginativa y detallada la experiencia musical de los ciudadanos de Leipzig es simplemente que deberíamos llevarnos a nosotros mismos a ese viaje imaginario, y noso-

tros sabemos demasiado. Pero si queremos, podemos hacer una lista de todas las diferencias entre nuestras disposiciones y conocimientos y las de los ciudadanos de Leipzig y, comparando las listas, podemos llegar a apreciar, *en cualquiera de sus detalles*, las diferencias entre lo que hubiera sido para ellos escuchar a Bach y lo que es para nosotros. Así, aunque lamentemos esa inaccesibilidad, cuando menos podremos comprenderla. No habría ningún misterio; solamente una experiencia que podría describirse con bastante precisión, pero que no podríamos disfrutar a menos que dedicáramos un esfuerzo ridículo en reconstruir nuestras estructuras de disposiciones personales.

Los qualófilos, sin embargo, han opuesto una fuerte resistencia a esta conclusión. Les parece que, aun en el caso de que una investigación como la que acabamos de imaginar pudiera establecer *casi* todos los puntos de lo que habrían sentido los ciudadanos de Leipzig, siempre quedaría un residuo inefable, *algo* acerca de lo que habrían sentido los ciudadanos de Leipzig que ningún nuevo avance en meros conocimientos «disposicionales» o «mecanicistas» podría reducir a cero. Éste es el motivo por el cual los qualófilos deben recurrir a los qualia en tanto que rasgos *adicionales*, además e independientemente de las configuraciones que determinan la retirada, el fruncir el ceño, el gritar y otras «meras conductas» de disgusto, odio o miedo. Podemos ver esto más claramente si volvemos a nuestro ejemplo de los colores.

Supongamos que le sugerimos a Otto que lo que hizo que su «rosa ocurrente» fuese esa experiencia tan irresistible de que disfrutó no es más que la suma total de todas las asociaciones y disposiciones reactivas innatas y aprendidas desencadenada por el modo en que sus ojos lo (mal) informaron:

Lo que *son* los qualia, Otto, no es más que esos complejos de disposiciones. Cuando usted dice «*Esto* es mi quale», lo que está usted identificando o a lo que está usted haciendo referencia, *tanto si se da cuenta como si no*, es a su complejo idiosincrásico de disposiciones. Usted parece estar haciendo referencia a algo privado e inefable en su ojo de la mente, un tono privado de rosa homogéneo, pero eso no es más que lo que a usted le parece, y no lo que es en realidad. Ese «quale» suyo es un personaje de pleno derecho en el mundo ficticio de su heterofenomenología, pero lo que resulta ser en el mundo *real* de su cerebro no es más que un complejo de disposiciones.

> «Eso no puede ser todo [replica Otto, dando el paso fatal dentro de la tradición qualófila], pues, mientras que un complejo de meras disposiciones podría ser la base o la fuente, de un modo u otro, de mi quale particular del rosa, todas estas disposiciones podrían verse alteradas sin que por ello se viera alterado mi quale intrínseco o, alternativamente, mi quale intrínseco podría cambiar sin cambiar la secuencia de meras disposiciones. Por ejemplo, mis qualia podrían *invertirse* sin que se invirtieran mis disposiciones. Podría tener todas las reactividades y asociaciones que ahora tengo para *verde* acompañadas por el *quale* que ahora tengo para *rojo*, y viceversa.»

4. Una fantasía filosófica: qualia invertidos

La idea de que sea posible la existencia de estos «qualia invertidos» es uno de los memas más virulentos de la filosofía. Locke la discutió en su *Ensayo sobre el entendimiento humano* (1690), y muchos de mis estudiantes me cuentan que de niños tuvieron la misma idea y que se sintieron fascinados por ella. La idea parece ser bastante clara:

> Está la manera en que yo veo, oigo, huelo, etc., las cosas. Todo esto es obvio. Me pregunto, sin embargo, si la manera en que las cosas me aparecen a mí es la misma en que les aparecen a los demás.

Los filósofos han compuesto numerosas variaciones sobre este tema, pero la versión clásica es la versión interpersonal: ¿Cómo sé que usted y yo vemos el mismo color *subjetivo* cuando miramos algo? Habida cuenta de que ambos aprendimos las palabras para los colores mostrándonos objetos coloreados, nuestras conductas verbales coincidirán aun en el caso de que experimentemos colores subjetivos completamente distintos; incluso si la manera en que yo veo el rojo es igual a la manera en que usted ve el verde, por ejemplo. En público, diremos de las mismas cosas que son «rojas» y «verdes», incluso si nuestras experiencias privadas son «las opuestas» (o diferentes).

¿Existe alguna manera de descubrir si es así? Considérese la hipótesis de que tanto usted como yo vemos igual las cosas rojas. ¿Es esta hipótesis irrefutable e inconfirmable a la vez? Muchos han pensado que sí, y algunos han llegado a la conclusión de que, precisamente por este motivo, se trata de un sinsentido, a pesar de su apelación inicial al sentido común. Otros se han preguntado si la tecnología podría acudir en nuestro rescate y confirmar (o desconfirmar) la hipótesis del espectro invertido interpersonal. La película de ciencia ficción *Brainstorm** (que no es, me apresuro a decir, una versión de mi libro *Brainstorms*) nos presenta el dispositivo imaginario adecuado: un aparato desarrollado con técnicas neurocientíficas se coloca sobre su cabeza y transmite sus experiencias visuales a mi cerebro a través de un cable. Con los ojos cerrados yo puedo referir con todo lujo de detalles lo que está usted mirando, con el añadido de que me maravillo de ver el cielo amarillo, la hierba roja, etc. Si dispusiéramos de una máquina como ésta, ¿no podría acaso un experimento realizado con ella confirmar, empíricamente, la hipótesis de que nuestros qualia son diferentes? Pero supongamos ahora que el técnico desenchufa el cable de conexión, lo invierte 180 grados, y lo vuelve a enchufar, y ahora yo refiero que el cielo es azul, la hierba verde, etc. ¿Cuál sería la orientación «correcta» del enchufe? Diseñar y construir un dispositivo como éste —en el supuesto de que fuera posible— requeriría que su «fidelidad» se ajustara o calibrara por la normalización de los testi-

* En España, la película se estrenó con el título de *Proyecto Brainstorm*. [N. del T.]

monios de ambos sujetos, de modo que nuestra evidencia volvería a estar en el punto de partida. Ahora podríamos intentar evitar esta conclusión elaborando argumentos adicionales, pero existe consenso entre los qualófilos de que ésta es una causa perdida; todo el mundo parece estar de acuerdo en que la moraleja de este experimento mental es que la comparación intersubjetiva de los qualia es imposible, incluso con una tecnología perfecta. Ello apoya, sin embargo, la visión «verificacionista» o «positivista» de que la idea de los qualia invertidos no tiene sentido y, en consecuencia, de que la idea misma de los qualia también carece de sentido. Como lo expresó el filósofo Ludwig Wittgenstein, utilizando su famosa analogía del «escarabajo en la caja»,

> La cosa que hay en la caja no pertenece en absoluto al juego del lenguaje; ni siquiera como un algo: pues la caja podría incluso estar vacía. —No se puede «cortar por lo sano» por la cosa que hay en la caja; se neutraliza, sea lo que fuere (1953, pág. 100).*

¿Pero qué significa esto? ¿Significa acaso que los qualia son reales pero inefectivos? ¿O que, después de todo, los qualia no existen? Que los qualia eran reales, todavía le parecía obvio a la mayoría de los filósofos que pensaban en ello, incluso si una diferencia en cuanto a los qualia fuese una diferencia que no pudiera ser detectada de ningún modo. Así quedaron las cosas, un tanto intranquilas, hasta que alguien imaginó lo que podría ser una versión mejorada del experimento mental: el espectro invertido *intra*personal. Parece ser que la misma idea se les ocurrió a varias personas independientemente (Gert, 1965; Putnam, 1965; Taylor, 1966; Shoemaker, 1969; Lycan, 1973). En esta versión, las experiencias que se comparan se producen todas en la misma mente, así que no tenemos que recurrir a la inútil máquina Brainstorm.

> Se levanta usted por la mañana y descubre que la hierba se ha vuelto roja, el cielo amarillo, etc. Nadie más nota esas anomalías cromáticas en el mundo, así que el problema debe de estar en usted. Parece que usted no tiene más remedio que concluir que ha sufrido una inversión de los qualia cromáticos. ¿Cómo ocurrió? Resulta que mientras dormía, unos neurocirujanos malvados cambiaron los cables —las neuronas— que conducen a las células cónicas sensibles al color de sus retinas.

Hasta aquí, estupendo. El efecto que esto tendría en usted sería sobrecogedor, puede que incluso terrorífico. Sin duda usted sería capaz de detectar que la manera en que ve las cosas es ahora muy diferente, e incluso dispondríamos de una explicación científica adecuada de por qué es así: los grupos de neuronas en el córtex visual que «se interesan» por el color, por ejemplo, estarían recibiendo sus estímulos de un conjunto sistemáticamente

* Página 245 de la edición bilingüe (Crítica, 1988). [N. del T.]

cambiado de receptores retinales. Así pues, parece que tenemos media batalla ganada: una diferencia en los qualia sería detectable, después de todo, si fuese una diferencia que se produjera de forma repentina en una misma persona.[4] Pero esto es sólo parte de la batalla, ya que la broma neuroquirúrgica que hemos imaginado también ha cambiado todas sus disposiciones reactivas; no sólo *dice* usted que sus experiencias cromáticas están todas liadas, sino que también ha sido invertida su conducta no verbal relacionada con el color. La irritabilidad que usted solía mostrar ante la luz roja, ahora se manifiesta en presencia de la luz verde, y usted ha perdido la fluidez con que solía utilizar los diversos esquemas de codificación por colores de su vida. (Si usted juega al baloncesto en los Boston Celtics, sigue pasando el balón por error a los que visten el uniforme *rojo*.)

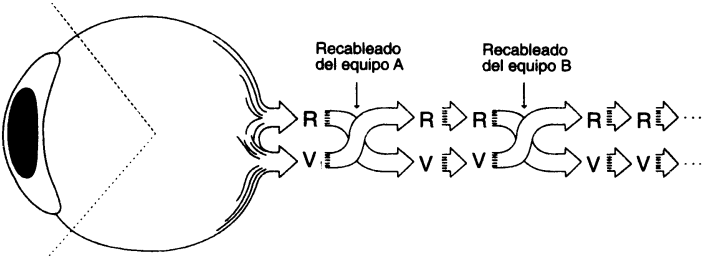


Figura 12.1

Lo que necesita el qualófilo es un experimento mental que demuestre que el modo en que las cosas se ven puede ser independiente de todas estas disposiciones reactivas. Así pues, tenemos que complicar la historia con algún nuevo desarrollo; tenemos que describir algún acontecimiento que deshaga el cambio en las disposiciones reactivas pero que mantenga intacto el cambio de los «qualia». En este punto es donde la bibliografía se sumerge en fantasías cada vez más delirantes, ya que nadie piensa de hecho que el modo en que se ven las cosas pueda estar en ningún momento *realmente* separado de las disposiciones reactivas del sujeto; lo que ocurre es que los qualófilos consideran que ésta es *en principio una posibilidad* importante. Para demostrarlo, necesitan describir un caso posible en el que, por extravagante que sea, resulte obvio que esta separación es real. Considérese una historia que *no* funcionaría:

4. El carácter repentino del fenómeno sería importante, ya que si se produjera de forma gradual, usted podría no llegar a notarlo. Como señala Hardin (1990), con la edad, nuestros cristalinos se van amarilleando, lo cual altera nuestra manera de percibir los colores primarios; si se nos muestra una rueda de colores y se nos pide que señalemos el rojo puro (rojo sin naranja ni violeta), el punto del *continuum* al que señalemos depende en parte de la edad.

> «Una noche, mientras usted duerme, unos perversos neurocirujanos le cambian todos los cables que vienen de las células cónicas (como antes), pero después, más tarde durante esa noche tan ajetreada, otro equipo de neurocirujanos, el equipo B, aparece y lleva a cabo un recableado *complementario* en el mismo nervio óptico, pero un poco más arriba.»

Esta segunda operación reinstaura las viejas disposiciones reactivas (es de suponer), pero, ay, también reinstaura los viejos qualia. Las células en el córtex que «se ocupan» de los colores, por ejemplo, recibirían ahora las mismas señales que antes del primer recableado, gracias a la pronta intervención del equipo B en deshacer el daño. El segundo recableado se produjo demasiado pronto, aparentemente; se produjo *en el camino hacia* la experiencia consciente. Así pues, tenemos que contar una historia diferente, en que el segundo recableado se produzca más tarde, *después* de que los qualia invertidos hayan salido a saludar en la conciencia, pero *antes* de que se inicie cualquiera de las reacciones invertidas a estos qualia. Pero, ¿es esto posible? No si los argumentos en favor del modelo de las Versiones Múltiples son correctos. No existe ninguna línea que pueda ser trazada en la «cadena» causal que, a través de la conciencia, va desde el globo ocular hasta la conducta subsiguiente, de modo que todas las reacciones a x se produzcan después y la conciencia de x no se produzca antes. El motivo es que no es una cadena causal, sino una red causal, con vías múltiples en las que, simultánea y semiindependientemente, se van corrigiendo Versiones Múltiples. La historia del qualófilo tendría sentido si hubiera un Teatro Cartesiano, un lugar especial en el cerebro donde se produce la experiencia consciente. Si existiera un lugar así, podríamos aislarlo entre dos recableados, dejando a los qualia invertidos en el Teatro, pero manteniendo las disposiciones reactivas normales. Toda vez que el Teatro Cartesiano no existe, el experimento mental no tiene ningún sentido. No hay ninguna manera coherente de contar la historia que necesitamos. No hay ninguna manera de aislar las propiedades *presentadas* en la conciencia de las múltiples reacciones del cerebro a sus propias discriminaciones, ya que no hay ningún proceso de presentación adicional.

En la bibliografía sobre el espectro invertido, a veces se supone que el segundo recableado se produce no por cirugía, sino por un proceso gradual de adaptación del sujeto a un nuevo régimen de experiencias. Esto tiene un cierto sentido; es sorprendente la facilidad con que las personas se adaptan a los más extravagantes desplazamientos de los sentidos. Se han llevado a cabo numerosos experimentos de inversión del campo visual en que los sujetos llevan unos anteojos que lo vuelven todo del revés, es decir, ¡ponen derecha la imagen retinal! (Por ejemplo, Stratton, 1896; Kohler, 1961. Welch, 1978, presenta un buen resumen de estos experimentos; véase también Cole, 1990.) Después de muchos días llevando unos anteojos de uno u otro tipo (hay diferencias, algunas variedades tienen un amplio campo de visión, otras

producen en los sujetos una especie de visión en túnel), los sujetos muestran a menudo un grado de adaptación sorprendente. En la película de Ivo Kohler sobre sus experimentos en Innsbruck, vemos a dos de sus sujetos cómicamente desamparados en el momento de ponerse los anteojos y, después, esquiando montaña abajo o circulando en bicicleta entre el tráfico de la ciudad, todavía llevando los anteojos y, en apariencia, completamente adaptados a ellos.

Supongamos, pues, que usted se adapta gradualmente a la inversión quirúrgica de su visión en color. (Por qué querría usted adaptarse, o tendría usted que adaptarse, ya es otro problema, pero podemos transigir en este punto ante los qualófilos, para precipitar su caída.) Así pues, al principio, algunas adaptaciones serían claramente *posteriores* a la experiencia. Podemos suponer que usted seguiría *viendo* amarillo el cielo, pero empezaría a llamarlo azul para no desentonar entre sus vecinos. Al mirar un objeto nuevo, se producirían unos breves instantes de confusión: «Es ver... ¡quiero decir *rojo!*». ¿Y qué decir de su irritabilidad ante la luz verde? ¿Seguiría apareciendo como una anormalidad en su respuesta galvánica cutánea? En aras de la coherencia argumental, el qualófilo tiene que imaginar, por inverosímil que parezca, que *todas* sus disposiciones reactivas se adaptan, dejando únicamente el residuo de los qualia invertidos, así que, de nuevo en aras de la coherencia argumental, aceptaremos que las predisposiciones innatas y más fundamentales de sus espacios cualitativos también «se adaptan»; esto es ridículo, pero lo peor está aún por venir.

A fin de poder contar la historia que necesita, el qualófilo tiene que suponer que finalmente todas estas adaptaciones se convierten en su segunda naturaleza, rápidamente y sin afectación. (Si no se convirtieran en su segunda naturaleza, entonces habría un remanente de disposiciones reactivas que seguirían siendo distintas, pero nuestra argumentación exige que no haya irregularidades.) Que así sea, pues. Ahora bien, en el supuesto de que *todas* sus disposiciones reactivas se reinstauren, ¿cuáles son sus propias intuiciones sobre sus qualia? ¿Están realmente invertidos o no lo están?

Llegados a este punto, me parece legítimo abandonar, habida cuenta de que después de haber tenido que tolerar tantos supuestos más que dudosos en aras de la coherencia argumentativa, uno o se queda vacío —ya no suben más burbujas de intuición— o ya recela de todas sus intuiciones. Pero quizá a usted le parezca obvio que sus qualia estuviesen invertidos. Aunque, ¿por qué? ¿Qué hay en la historia que le permita entenderlo así? Quizás, pese a haber seguido las instrucciones, usted ha añadido algún supuesto de más, o no ha sabido ver ciertas posibilidades que la historia no prohíbe. Creo que la explicación más razonable del hecho de que, en este caso imaginario, usted mantenga la intuición de que seguiría teniendo «qualia invertidos» es que usted está suponiendo, de forma ilegítima, que la adaptación se produce «en el momento posterior a la experiencia».

¿No podría darse el caso, quizá, de que la adaptación se produjera en la vía hacia la experiencia? Cuando nos ponemos unas gafas con los cristales

tintados, no vemos ningún color —o los colores que vemos son raros y difíciles de distinguir—, pero al cabo de un tiempo de llevarlas vuelve, sorprendentemente, la visión en color normal. (Cole, 1990, llama la atención de los filósofos sobre este punto, que usted mismo puede verificar con unas gafas de visión infrarroja adquiridas en una tienda de material militar de desecho.) Quizá, por desconocer este hecho sorprendente, a usted nunca se le ocurrió que *podría* adaptarse a la cirugía exactamente de la misma manera. Podríamos haber resaltado este punto en el experimento mental, añadiendo algunos detalles:

> «...Y, a medida que se iba produciendo la adaptación, usted iba notando, para su sorpresa, que los colores de las cosas no eran tan extraños después de todo, y, en ocasiones, usted se confundía y hacía correcciones *dobles*. Cuando le preguntaban por el color de un objeto nuevo, usted decía, "es ver... no es rojo... ¡no!, *¡es* verde!".»

Contada así, la historia haría parecer «obvio» que los qualia cromáticos se han adaptado o reinvertido. Pero en cualquier caso, pensará usted ahora, tiene que ser una cosa u otra. ¡No es posible hallar un caso en que el tipo de reajuste producido no fuese obvio! El supuesto sobre el que se basa *esta* convicción es que todas las adaptaciones pueden categorizarse como previas o posteriores a la experiencia (estalinianas u orwellianas). Al principio podría parecer un supuesto bastante inocente, ya que los casos extremos son fáciles de clasificar. Cuando el cerebro introduce compensaciones a los movimientos de los ojos y de la cabeza, produciendo un mundo visual estable «en la experiencia», está claro que nos hallamos ante una anulación de los efectos previa a la experiencia, una adaptación en la vía hacia la conciencia. Y cuando usted se imagina introduciendo compensaciones («tardías») en la elección de palabras que denotan colores («es ver..., ¡quiero decir *rojo!*»), está claro que estamos ante un ajuste posterior a la experiencia, meramente comportamental. Parece lógico, pues, que cuando se han producido *todas* las adaptaciones, o bien el color subjetivo (el color «en la conciencia») permanece invertido o bien no permanece invertido, ¿no? Así es como podríamos determinarlo: añádanse los recableados en la vía hacia la conciencia; si su número de orden es par —como en el caso de la manipulación efectuada por el equipo B—, entonces los qualia se han normalizado; si es impar, los qualia están invertidos. No tiene sentido. Recuérdese la curva neo-Laffer del capítulo 5. El hecho de que exista un único valor para una variable discriminada que pueda ser identificado como *el* valor de esa variable «en la conciencia» no es una necesidad ni lógica ni geométrica.

Podemos demostrarlo con una fantasía de cosecha propia, jugando con las reglas del qualófilo. Supongamos que, antes de la cirugía, un determinado tono de azul tiende a recordarnos un coche con el que sufrimos un accidente y que, por tanto, es un color que tendemos a rechazar. Al principio, después de la cirugía, no mostramos ninguna reacción negativa ante las co-

sas que son de este color, pues lo vemos, por ejemplo, como un amarillo inocuo y falto de todo poder para hacernos recordar. Una vez completado nuestro proceso de adaptación, no obstante, volvemos a rechazar ese tono de azul, *porque nos recuerda el accidente.* (Si no lo hiciera, estaríamos ante una disposición reactiva no adaptada.) Pero si nos preguntaran si ello se debe a que, tal como recordamos el accidente, el coche era amarillo —exactamente del mismo color que ese objeto horrible que ahora está ante nosotros— o se debe a que, tal como recordamos el accidente, el coche era azul —exactamente del mismo color que ese objeto horrible que ahora está ante nosotros—, no seríamos capaces de responder. Nuestra conducta verbal estaría completamente «adaptada»; nuestra respuesta inmediata, fruto de nuestra segunda naturaleza, a la pregunta «¿de qué color era el coche con que tuvo el accidente?» sería «azul», y, sin dudarlo, también diríamos que el objeto horrible que está ante nosotros es azul. ¿Significa esto que hemos *olvidado* el largo período de entrenamiento?

No. No necesitamos algo tan dramático como la amnesia para explicar nuestra incapacidad para responder, ya que disponemos de muchos casos cotidianos en los que se produce el mismo fenómeno. ¿Le gusta la cerveza? Muchas personas a las que les gusta la cerveza reconocerán que la cerveza es un gusto adquirido. Poco a poco, uno se acostumbra —o llega— a disfrutar de su sabor. ¿Qué sabor? ¿El sabor del primer sorbo?

> «A nadie podría gustarle ese sabor [replicaría un experimentado bebedor de cerveza]. La cerveza les sabe diferente a los bebedores experimentados. Si la cerveza me siguiera sabiendo igual que como me supo el primer sorbo, ¡nunca hubiera seguido bebiendo cerveza! O, por plantear la misma cuestión desde el otro punto de vista, si el primer sorbo de cerveza me hubiera sabido de la manera que me sabe ahora, ¡nunca habría tenido que adquirir el gusto! Me habría gustado el primer sorbo tanto como me gusta el de la cerveza que me estoy tomando.»

Si nuestro bebedor de cerveza está en lo cierto, entonces la cerveza *no* es un gusto adquirido. Nadie llega a disfrutar *de la manera que sabe la cerveza la primera vez que se prueba.* Por el contrario, *la manera como nos sabe la cerveza* cambia gradualmente. Otros bebedores de cerveza podrían insistir en que, de ningún modo, la cerveza *siempre* les ha sabido como al principio, sólo que ahora les gusta *ese sabor.* ¿Existe alguna diferencia? Evidentemente, existe una diferencia en el nivel de la heterofenomenología, y esa diferencia debe ser explicada. Podría ser que esas convicciones diferentes surgiesen a partir de genuinas diferencias en la capacidad discriminativa como las siguientes: en el primer tipo de bebedor de cerveza, la «formación» ha alterado la «forma» del espacio cualitativo de degustación, mientras que en el segundo caso, el espacio cualitativo apenas se ve alterado, mientras que la «función de evaluación» sobre ese espacio ha sufrido un proceso de revisión. O también *podría* ser que los bebedores de cerveza se estuvieran enga-

ñando a sí mismos (como los que insisten en decir que las Marilyns de alta resolución están realmente ahí, en un segundo plano de su campo visual). Tenemos que mirar más allá de los mundos heterofenomenológicos hacia los acontecimientos reales que se producen dentro de nuestras cabezas a fin de comprobar si existe una interpretación, *salva veritate* (por «forzada» que sea) de las afirmaciones de los bebedores de cerveza, y si existe, será solamente porque habremos decidido *reducir* «la manera como sabe» a un complejo u otro de disposiciones reactivas (Dennett, 1988a). Tendremos que destruir los «qualia» para poder «salvarlos».

Así pues, si un bebedor de cerveza arruga la frente, hace una mueca de profunda seriedad y dice que aquello a lo que se está refiriendo es «*la manera como me sabe la cerveza en este momento*», entonces ya no cabe duda de que se está engañando a sí mismo si piensa que *así* puede hacer referencia a un quale de su conocimiento, un estado subjetivo que es independiente de sus actitudes reactivas cambiantes. Quizá le parezca que puede, pero no puede.[5]

Por la misma razón, en el caso imaginario en que el objeto azul nos recordaba el accidente de coche, nos estaríamos engañando a nosotros mismos si pensáramos que somos capaces de distinguir, a partir de *cómo vemos los objetos*, si es «intrínsecamente» igual a cómo veíamos el coche en el momento del accidente. Con esto basta para abaratar el experimento mental del qualófilo, ya que el objetivo consistía en describir un caso en el que fuera *obvio* que los qualia estarían invertidos mientras que las disposiciones reactivas eran normalizadas. El *supuesto* de que uno es capaz de distinguir constituye una petición de principio y, sin supuesto no hay argumento, sólo hay una bomba de intuiciones, una historia que le engatusa para que declare cuáles son sus intuiciones más privadas, aunque sin darle una buena razón para ello.

Con petición de principio o sin ella, aún puede parecer obvio que «los colores subjetivos que vemos en las cosas» *deberían* ser «de una manera o de otra». Todo lo cual no es más que una demostración de la poderosa fuerza gravitatoria que ejerce el Teatro Cartesiano sobre nuestra imaginación. A fin de acabar de una vez por todas con el atractivo que todavía pudiera tener esta idea, podría sernos útil volver a considerar el caso paralelo de los anteojos que invierten la imagen. Cuando las adaptaciones de los sujetos que llevan dichos anteojos se han convertido en una segunda naturaleza, entonces pueden ir en bicicleta y esquiar, de modo que, de forma natural (pero equivocada), viene a la mente la siguiente pregunta: ¿se han adaptado *volviendo a poner derecho su mundo de las experiencias, o acostumbrándose a que su mundo de las experiencias esté al revés*? ¿Qué dicen ellos? Dicen cosas diferentes, que están en relación más o menos directa con el grado de

5. «El solo hecho de que queramos decir "lo importante es *esto*" —indicando la sensación para nosotros mismos— muestra ya cuán inclinados estamos a decir algo que no constituye información alguna.» Wittgenstein (1953), §298. [De la edición bilingüe (Crítica, 1986); N. del T.]

adaptación que han conseguido. Cuanto mayor es el grado de adaptación, más común es que los sujetos rechacen la pregunta como impropia o imposible de responder. Esto es precisamente lo que reclama la teoría de las Versiones Múltiples: toda vez que, distribuidas por el cerebro, tenemos una serie de discriminaciones y reacciones que deben ajustarse, algunas de ellas relacionadas con «reflejos» de bajo nivel (como apartarse en la dirección adecuada cuando algo se nos acerca) y otras relacionadas con acciones deliberadas que son el foco de nuestra atención, no debe sorprendernos que, a medida que se acumulan las adaptaciones en este mosaico, los sujetos pierdan toda convicción sobre si decir «veo las cosas como las veía antes» en vez de «lo veo todo diferente, pero me estoy acostumbrando». En algunos casos, ven las cosas iguales (a juzgar por sus reacciones), en otros las ven diferentes (a juzgar por sus reacciones). Si sólo hubiera una única representación del espacio visuomotor a través del cual tuvieran que canalizarse todas las reacciones a los estímulos visuales, *ésta* debería ser «de una manera o de otra», quizá, pero no sólo existe esa única representación. *La* manera como ven las cosas se compone de muchos hábitos de reacción parcialmente independientes, y no de una única imagen en la cabeza que esté intrínsecamente derecha o invertida. Lo único que importa es que la información de entrada y la información de salida encajen y, dado que ello se produce en muchos lugares y a través de medios distintos y, en gran medida, independientes, no se puede decir qué «equivale» a «mi campo visual sigue estando al revés».

Lo mismo es válido para la inversión de los «qualia». La idea de que se trata de algo *además* de la inversión de todas las disposiciones reactivas de uno, de modo que, si éstas se normalizaran de nuevo, los qualia permanecerían invertidos, no es más que parte del tenaz mito del Teatro Cartesiano. Este mito se solemniza en los complejos experimentos mentales sobre el espectro invertido, pero solemnizar no es evidenciar o probar. *Si* no hay qualia además de la suma total de disposiciones reactivas, la idea de mantener constantes los qualia mientras se reajustan las disposiciones es contradictoria en sí misma.

5. ¿Qualia «epifenoménicos»?

Existe otro experimento mental filosófico sobre nuestra experiencia cromática que se ha mostrado irresistible: el ampliamente discutido caso, expuesto por Frank Jackson (1982), de Mary, la investigadora del color que nunca había visto los colores. Como todo buen experimento mental, aquello que pretende demostrarnos es evidente incluso para el profano. De hecho, es un mal experimento mental, una bomba de intuiciones que nos anima a malinterpretar sus premisas.

> Mary es una brillante investigadora que, por los motivos que sea, se ha visto obligada a investigar el mundo desde una habitación en blanco y negro *a*

> *través* de un monitor de televisión en blanco y negro. Se especializó en la neurofisiología de la visión y llegó a adquirir, supongamos, toda la información física que se puede obtener sobre lo que ocurre cuando vemos unos tomates maduros o el cielo, y utilizamos términos como *rojo, azul*, etc. Descubre, por ejemplo, exactamente qué combinación de longitudes de onda procedentes del cielo estimulan la retina, y de qué modo ello produce, *a través* del sistema nervioso central, la contracción de las cuerdas vocales y la expulsión de aire de los pulmones que resulta en la proferencia del enunciado «el cielo es azul» (...) ¿Qué ocurriría si Mary fuese liberada de su encierro en blanco y negro o recibiera un monitor en color? ¿*Aprendería* algo o no? Resulta obvio que sí que aprenderá algo sobre el mundo y nuestra experiencia visual de éste. Pero, en ese caso, es ineludible la conclusión de que sus conocimientos previos eran incompletos. Y sin embargo, Mary poseía *toda* la información física. *Ergo*, se necesita algo más que eso, y el fisicalismo es falso (pág. 128).

Las cosas no podrían estar más claras. Mary *no* poseía ninguna experiencia previa del color (en la habitación no hay espejos donde pudiera mirarse, está obligada a llevar guantes negros, etc., etc.), de modo que, en el momento tan especial en que sus secuestradores la ponen en libertad en un mundo lleno de colores que sólo conoce a través de descripciones (y diagramas en blanco y negro), «resulta obvio», como dice Jackson, que aprenderá algo. En efecto, todos podemos imaginárnosla claramente mirando una rosa roja por primera vez y exclamando, «¡de modo que *así* es como se ve el rojo!». Y también podríamos pensar que si las primeras cosas coloreadas que ve son, por ejemplo, unos bloques de madera sin ninguna indicación sobre su color, y se le informa de que uno de ellos es rojo y el otro es azul, no tendrá ni la más remota idea de cuál es cuál hasta que no tenga la oportunidad de aprender en experiencias futuras qué palabras corresponden a cada color.

Así es como casi todo el mundo imagina este experimento mental; no sólo los profanos, sino también los filósofos más sagaces y duros de pelar (Tye, 1986; Lewis, 1988; Loar, 1990; Lycan, 1990; Nemirow, 1990; Harman, 1990; Block, 1990; van Gulick, 1990). Sólo Paul Churchland (1985, 1990) ha mostrado cierta resistencia a la *imagen*, que tan claramente despierta en nosotros el experimento mental, del dramático descubrimiento de Mary. La imagen es incorrecta; si es así como está usted imaginando el caso, entonces es que no está usted siguiendo las instrucciones. El motivo por el cual nadie sigue las instrucciones es porque lo que le piden que imagine es tan absurdamente inmenso, que ni siquiera puede intentarlo. La premisa clave es que «ella posee *toda* la información física». Esto no es fácil de imaginar, así que nadie se molesta en hacerlo. Todos se limitan a imaginar que sabe mucho, quizás incluso lleguen a imaginarse que sabe todo lo que *hoy* se sabe sobre la neurofisiología de la visión del color. Pero eso no es más que una gota de agua en el mar, y no debe sorprendernos que Mary aprendiera algo si *eso* fuese todo lo que sabe.

A fin de recalcar la ilusión de que es presa nuestra imaginación, permítaseme continuar la historia de una manera sorprendente, pero legítima:

> Y así, un buen día, los secuestradores de Mary decidieron que había llegado el momento de que pudiera ver los colores. Para engañarla, prepararon un plátano de color azul brillante para presentárselo como su primera experiencia cromática. Mary lo miró y dijo: «¡Hey! ¡Me habéis querido engañar! ¡Los plátanos son amarillos, pero éste es azul!». Los secuestradores se quedaron boquiabiertos. ¿Cómo lo hizo? «Fácil», replicó ella, «no tenéis que olvidar que yo sé todo —absolutamente todo— lo que puede saberse sobre las causas y efectos físicos de la visión en color. Así que antes de que me trajerais el plátano, yo ya tenía anotado, con todo lujo de detalles, qué impresión física exacta produciría un objeto amarillo o un objeto azul (o un objeto verde, etc.) en mi sistema nervioso. Así que yo ya conocía de antemano qué *pensamientos* tendría (porque, después de todo, la «mera disposición» a pensar sobre esto o aquello no es uno de sus famosos qualia, ¿no?). Mi experiencia del azul no (lo que me sorprendió es que intentarais pillarme con un truco tan barato). Me doy cuenta de que os *resulta muy difícil imaginar* que yo pueda saber tanto sobre mis disposiciones reactivas, hasta el punto de que el modo en que me afectó el azul no me sorprendiera. Claro que os resulta difícil. ¡Para todo el mundo es difícil imaginar las consecuencias de que alguien lo sepa absolutamente todo, sobre todo lo físico!

«Seguro que ha hecho trampas», pensará usted. Debo de haber ocultado alguna imposibilidad tras el velo de las observaciones de Mary. ¿Puede usted probarlo? Mi objetivo no es demostrar que mi manera de contar el resto de la historia prueba que Mary *no* aprende nada, sino que la manera tradicional de imaginar la historia no *prueba* que lo *haga*. No prueba nada; simplemente bombea la intuición de que es así («resulta obvio»), cautivándole y llevándole a imaginar una cosa distinta de lo que las premisas proponen.

Sin duda es cierto que en cualquier versión realista y fácil de imaginar de la historia, Mary llegaría a aprender algo, pero en cualquier versión realista y fácil de imaginar, ella podría saber mucho, pero no lo sabría todo sobre lo físico. El simple acto de imaginar que Mary sabe mucho, y no ir más allá, no es una buena manera de hacerse una idea de las implicaciones de que posea «toda la información física», del mismo modo que imaginar que es asquerosamente rica no es una buena manera de hacerse una idea de las implicaciones de la hipótesis de que lo posee todo. A fin de comprender mejor hasta qué punto sus conocimientos constituyen un poder, nos puede servir de ayuda hacer una lista de algunas de las cosas que conocería. Conoce el blanco y el negro, y los diferentes tonos de gris; también conoce la diferencia entre el color de un objeto y propiedades de las superficies tales como el ser brillante o el ser mate; y también conoce la diferencia entre fronteras de luminancia y fronteras de color (las fronteras de luminancia son las que vemos en los televisores en blanco y negro, por expresarlo de un modo no muy preciso). También conoce exactamente qué efectos —descritos en términos neurofisiológicos— tendría cada color sobre su sistema nervioso. Así que lo único que le queda por saber es una manera de identificar esos efectos neurofisiológicos «desde dentro». Supongo que usted se imaginará fácilmente a Mary realizando *pocos* progresos en este campo; por ejemplo,

buscando hábiles estrategias por las cuales sería capaz de decidir si un determinado color, sea cual sea, *no* es amarillo o no es rojo. ¿Cómo? Observando alguna reacción sobresaliente y específica que tendría su cerebro sólo ante el amarillo o sólo ante el rojo. Pero si le permitimos aunque sea ese pequeño acceso a su espacio cromático, entonces no tendremos más remedio que concluir que puede hallar un impulso para completar sus conocimientos, porque ella no conoce solamente las reacciones más *sobresalientes*, las conoce todas.

Recordemos la caja de gelatina que Julius y Ethel Rosenberg convirtieron en un detector-*M*. Imagine ahora su sorpresa si apareciera un impostor con un trozo «que coincide» pero no es el original. «¡Imposible!», exclamarían. «No, no es imposible», diría el impostor, «sólo es muy difícil. Yo disponía de *toda la información* necesaria para reconstruir un detector-*M* y para fabricar otra cosa con la propiedad formal *M*». Mary disponía de información suficiente (en el caso original, si imaginamos lo que hay que imaginar) para hacerse una idea de cuáles eran sus detectores de rojo y cuáles sus detectores de azul y poder, así, identificarlos de antemano. No es la manera habitual de aprender sobre los colores, pero Mary no es el tipo normal de persona.

Sé que esto no satisfará a muchos de los fans filosóficos de Mary, y que quedan todavía muchas cosas que decir, pero —y éste es mi objetivo principal— el acto real de probar debe llevarse a un terreno muy alejado del ejemplo de Jackson, que es un clásico provocador del «síndrome de los filósofos»: confundir una falta de imaginación por la intuición de una necesidad. Puede que a algunos filósofos que han tratado el caso de Mary no les importe haber imaginado mal el ejemplo, ya que lo han utilizado como trampolín en discusiones que iluminaban otras cuestiones interesantes e importantes, pero independientes. No me ocuparé de estas cuestiones aquí, ya que lo que me interesa es considerar la conclusión a que llega Jackson a partir de su ejemplo: las experiencias visuales tienen qualia que son «epifenoménicos».

El término «epifenómeno» es de uso común hoy en día, tanto en filosofía como en psicología (y en otras ciencias cognitivas). Se utiliza con la presunción de que su significado es conocido y acordado, cuando, de hecho, los filósofos y los investigadores en ciencia cognitiva lo utilizan con significados *completamente* diferentes; un hecho curioso, que resulta aún más extraño por el hecho de que, aunque yo me he dedicado a hacerlo notar una y otra vez, a nadie parece importarle. Puesto que el «epifenomenismo» a menudo parece ser el último refugio para los qualia, y ya que esta aparente seguridad se debe enteramente a la confusión entre ambos significados, me veo obligado a tomar cartas en el asunto y a poner en guardia a aquellos que utilizan el término.

Según el *Shorter Oxford English Dictionary*, el término «epifenómeno» aparece por primera vez en 1706 como un término de la patología, «un sín-

toma o aparición secundario».* El biólogo evolucionista Thomas Huxley (1874) es probablemente el escritor que amplió el significado del término en lo que es su uso más común en la psicología, donde significa propiedad *no funcional* o producto derivado. Huxley utilizó el término en su discusión de la evolución de la conciencia y en su afirmación de que las propiedades epifenoménicas (como el «silbido de un motor de vapor») no pueden ser explicadas por selección natural.

He aquí un caso claro de este uso de la palabra:

> ¿Por qué las personas que están pensando se muerden los labios o dan golpecitos con los pies? ¿Son estas acciones meros epifenómenos que acompañan a los procesos centrales del sentir y del pensar o podrían ser partes integrantes de estos procesos? (Zajonc y Markus, 1984, pág. 74).

Nótese que lo que los autores quieren decir es que estas acciones, a pesar de ser fácilmente detectables, no juegan ningún papel operativo ni ningún papel estructural en los procesos del sentimiento y del pensamiento; son no funcionales. En la misma línea, podemos decir que el ronroneo del ordenador o nuestra sombra cuando nos estamos haciendo una taza de té son epifenoménicos. Los epifenómenos son meros subproductos, pero como tales son productos con numerosos efectos en el mundo: dar golpecitos con los pies produce un ruido que se puede grabar, y nuestra sombra tiene efectos sobre el papel fotográfico, por no mencionar el ligero enfriamiento de las superficies que causa cuando se proyecta sobre ellas.

El significado filosófico tradicional del término es diferente: «x es epifenoménico» significa que «x es un efecto, pero que por sí mismo no tiene ningún tipo de efecto sobre el mundo físico». (Véase Broad, 1925, pág. 118, para una definición que inaugura, o en todo caso establece, el uso filosófico de la palabra.) ¿Son tan diferentes estos significados? Sí, tan diferentes como los significados de *asesinato* y *muerte*. El significado filosófico es más fuerte: una cosa que no tenga ningún efecto sobre el mundo físico no tendrá ningún efecto sobre la función de nada, pero lo inverso no *es* cierto, como evidencia el ejemplo de Zajonc y Markus.

De hecho, el significado filosófico es demasiado fuerte; da lugar a un concepto completamente inútil (Harman, 1990; Fox, 1989). Si x no tiene efectos físicos (de acuerdo con esta definición), entonces ningún instrumento podrá detectar la presencia de x ni directa ni indirectamente; el funcionamiento del mundo no se ve afectado en lo más mínimo por la presencia o la ausencia de x. ¿Cómo es posible, pues, que exista evidencia empírica alguna que nos permita afirmar la presencia de x? Supongamos, por ejemplo, que Otto insiste en afirmar que él (cuando menos) posee qualia epifenoménicos. ¿Por

* El *Diccionario de la lengua española* de la Real Academia no recoge el término en su edición de 1984, aunque sí aparece en el *Diccionario de filosofía* de J. Ferrater Mora con un significado muy próximo al que Dennett presenta más adelante. [N. del T.]

qué querría decir algo así? No será porque éstos tienen algún efecto sobre él, guiándolo y alertándolo a medida que hace sus declaraciones. Por la definición misma de epifenómeno (en el sentido filosófico), las sinceras declaraciones de Otto de que posee epifenómenos *no podrían* constituir evidencia suficiente para él mismo o para cualquier otro de que realmente los posee, ya que podría estar diciendo exactamente lo mismo si no los tuviera. Pero, quizás Otto disponga de alguna evidencia «interna».

Aquí tenemos una posible escapatoria, aunque no muy atractiva. Los epifenómenos, recuerde, se definen por no tener ningún efecto sobre el mundo *físico*. Si Otto está dispuesto a abrazar un dualismo acérrimo, puede afirmar que sus qualia epifenoménicos no tienen efectos sobre el mundo físico, pero sí tienen efectos sobre su mundo mental (no físico); Broad (1925), cerró esta vía de escape por definición, pero uno es libre de volver a abrirla. Por ejemplo, éstos *causan algunas de sus creencias (no físicas)*, como por ejemplo la creencia de que posee qualia epifenoménicos. Pero ésta no es más que una vía de escape que lo saca temporalmente del azoramiento. Pues ahora, si no quiere caer en una enorme contradicción, debe aceptar que sus creencias tampoco tienen ningún efecto sobre el mundo físico. Si de repente llegara a perder sus qualia epifenoménicos, dejaría de creer que los tuvo en algún momento, aunque seguiría *diciendo* que los tuvo. ¡No creería en lo que dice! (Ni tampoco podría decirnos que no creía en lo que estaba diciendo, ni podría hacer nada que revelara que ya no creía en lo que estaba diciendo.) Así pues, la única manera en que Otto puede «justificar» su creencia en epifenómenos consistiría en retirarse a un mundo solipsístico cuyos únicos habitantes son él, sus creencias y sus qualia, separados de todos los efectos sobre el mundo. Lejos de ser una manera «segura» de ser materialista y de guardar los qualia, esto es a lo sumo una manera de abrazar el más radical de los solipsismos, aislando nuestra mente —nuestras creencias y nuestras experiencias— de cualquier comercio con el mundo material.

Si los qualia son epifenoménicos en el sentido filosófico tradicional, su existencia no puede explicar cómo ocurren las cosas (en el mundo material), ya que, por definición, las cosas ocurrirían igual sin ellos. No puede haber, pues, ninguna razón empírica para creer en epifenómenos. ¿Podría haber algún otro tipo de razón para afirmar su existencia? ¿Qué tipo de razón? Una razón *a priori*, presumiblemente. ¿Pero cuál? Nadie ha presentado nunca ninguna —buena, mala o indiferente— que yo sepa. Si alguien quisiera objetar que estoy actuando como un «verificacionista» en este asunto de los epifenómenos, mi réplica no puede ser más que la siguiente: ¿Acaso no es todo el mundo verificacionista ante *este* tipo de afirmaciones? Considérese, por ejemplo, la hipótesis de que en cada cilindro de todo motor de explosión hay catorce *gremlins* epifenoménicos. Estos *gremlins* carecen de masa, de energía y de propiedades físicas; no hacen que el motor funcione con más o menos suavidad, con más o menos rapidez. No hay *ni puede haber* nada que evidencie su presencia, ni hay tampoco ningún medio empírico que nos permita distinguir esta hipótesis de sus rivales: los *gremlins* son doce o trece

o quince o... ¿A qué principio puede uno apelar para defender el rechazo total de un sinsentido como éste? ¿A un principio verificacionista, o simplemente al sentido común?

> «¡Ah, pero hay una diferencia [dice Otto.] No hay ninguna motivación independiente que nos permita tomarnos en serio su hipótesis de los *gremlins*. Usted acaba de inventársela para la ocasión. Los qualia, por el contrario, tienen una larga tradición y juegan un papel fundamental en nuestro esquema conceptual.»

¿Y qué diríamos si unos ignorantes que durante generaciones han venido creyendo que eran *gremlins* lo que hacía funcionar sus coches, ahora se han visto obligados por el empuje de la ciencia a sostener la afirmación desesperada de que los *gremlins* están ahí, pero son epifenoménicos? ¿Sería un error por nuestra parte el rechazar esta «hipótesis» sin más? Sea cual sea el principio en que nos basemos cuando le damos la espalda a este sinsentido, basta para rechazar la doctrina de que los qualia son epifenoménicos en el sentido filosófico. No son éstas ideas que uno deba discutir con una cara seria.

Resulta difícil creer que los filósofos que recientemente han calificado sus visiones como epifenomenismos puedan estar cometiendo un error tan desolador. ¿Acaso están afirmando que los qualia son epifenoménicos en el sentido de Huxley? En este sentido, los qualia *son* efectos físicos y *tienen* efectos físicos, aunque son no funcionales. Cualquier materialista estaría dispuesto a aceptar esta hipótesis como cierta, si identificamos los qualia con disposiciones reactivas, por ejemplo. Como observamos en nuestro examen del acto de disfrutar, aunque algunas deformaciones o predisposiciones de nuestros espacios cualitativos son funcionales —o solían ser funcionales—, otras son meras casualidades. ¿Por qué no me gusta el brécol? Probablemente por ningún motivo en concreto; mi disposición reactiva negativa es puramente epifenoménica, un subproducto de mi configuración sin ningún significado. No tiene función, pero tiene muchos efectos. En cualquier sistema diseñado, algunas propiedades son cruciales, mientras que otras son más o menos revisables *ad libitum*. Todo tiene que ser de una manera o de otra, pero a menudo los modos no importan. La palanca de cambio de un coche tiene que ser de una longitud determinada y de una fuerza determinada, pero si su corte transversal es circular, cuadrado u ovalado es una propiedad epifenoménica, en el sentido de Huxley. En los sistemas CADBLIND que imaginamos en el capítulo 10, el esquema de código de colores era epifenoménico. Podríamos «invertirlo» (utilizando números negativos, o multiplicando todos los valores por alguna constante) sin que haya ninguna diferencia *funcional* en sus capacidades de procesamiento de la información. Dicha inversión podría no ser detectable en una inspección causal, y podría no ser detectable *por el sistema*, pero no sería epifenoménica en el

sentido filosófico. Habría muchas pequeñas diferencias de voltaje en los registros de memoria donde se almacenan los números, por ejemplo.

Si pensamos en todas las propiedades de nuestro sistema nervioso que nos permiten ver, oír, oler, saborear y tocar las cosas, veremos que las podemos dividir, aproximadamente, en propiedades que juegan papeles realmente cruciales en mediar el procesamiento de la información, y en propiedades epifenoménicas que son más o menos revisables *ad libitum*, como el sistema de código de colores en el sistema CADBLIND. Cuando un filósofo hace la conjetura de que los qualia son propiedades epifenoménicas de los estados cerebrales, ello podría significar que los qualia podrían ser variaciones locales en el calor generado por el metabolismo neuronal. Esto no puede ser lo que los epifenomenistas tienen en mente, ¿no? Si lo es, entonces los qualia como epifenómenos no constituyen una amenaza para el materialismo.

Ha llegado el momento de traspasar el peso de la prueba a aquellos que persisten en utilizar el término. El sentido filosófico de este término es, simplemente, ridículo; el sentido de Huxley es relativamente claro y libre de problemas, pero también irrelevante para la argumentación filosófica. No hay ningún otro sentido del término que tenga algún valor. Así pues, si alguna vez alguien afirma que sostiene alguna variedad de epifenomenismo, sea educado, pero pregúntele: *¿de qué* está usted hablando?

Nótese, por cierto, que este equívoco entre los dos sentidos de «epifenoménico» también contamina los debates sobre el problema de los zombíes. Un zombie de filósofo, recuerde, desde el punto de vista de la conducta no es distinguible de un ser humano, pero no es consciente. No hay nada que se parezca al hecho de ser como un zombie; sólo les parece que sea así a los observadores (incluido él mismo, como vimos en el capítulo anterior). Esto puede tener una interpretación fuerte o una interpretación débil, dependiendo de cómo queramos tratar dicha indistinguibilidad de los observadores. Si quisiéramos afirmar que, *en principio*, un zombie no se puede distinguir de una persona consciente, entonces estaríamos diciendo que la conciencia genuina es epifenoménica *en el sentido ridículo del término*. Esto es una tontería. Así que podríamos decir, por el contrario, que la conciencia podría ser epifenoménica en el sentido de Huxley: aunque hubiera alguna manera de distinguir a los zombíes de las personas reales (quién sabe, a lo mejor los zombíes tienen el cerebro verde), tal diferencia no aparece *a los observadores* como una diferencia funcional. De forma equivalente, los cuerpos humanos con el cerebro verde no ocultan observadores, mientras que otros cuerpos humanos sí. Gracias a esta hipótesis, seríamos capaces, en principio, de distinguir los cuerpos habitados de los cuerpos deshabitados con sólo mirar el color del cerebro. Esto también es una tontería, evidentemente, y muy peligrosa, pues recuerda el tipo de prejuicios totalmente inmotivados que negaron el carácter de persona a ciertos seres humanos en función del color de su piel. Ha llegado el momento de calificar la idea de que puedan existir los zombíes como lo que es: una idea filosófica poco seria y una reliquia ridícula e innoble fruto de viejos prejuicios. ¡A lo mejor las muje-

res no son conscientes! ¡O los judíos! ¡Qué sinsentido tan pernicioso! Como dice Shylock, llamando la atención, con mucho tino, sobre criterios «meramente comportamentales»:

> ¿No tiene ojos un judío? ¿No tiene manos un judío, ni órganos, proporciones, sentidos, pasiones, emociones? ¿No toma el mismo alimento, le hieren las mismas armas, le atacan las mismas enfermedades, se cura por los mismos métodos? ¿No le calienta el mismo estío que a un cristiano? ¿No le enfría el mismo invierno? ¿Es que no sangramos si nos espolean? ¿No reímos si nos hacen cosquillas? ¿No morimos si nos envenenan?*

Existe otra manera de enfrentarse a la posibilidad de que haya zombíes que, en cierto modo, considero que es más satisfactoria. ¿Son posibles los zombíes? No sólo son posibles, son reales. Todos somos zombíes.[6] Nadie es consciente, en cualquier caso, ¡no en el modo sistemáticamente misterioso que apoya doctrinas tales como el epifenomenalismo! No puedo probar que este tipo de conciencia no exista. Tampoco puedo probar que los *gremlins* no existan. Lo mejor que puedo hacer es demostrar que no existe ningún motivo respetable para creer en ella.

6. De vuelta a mi mecedora

En el capítulo 2, apartado 2, inicié mi tarea de explicar la conciencia, recordando un episodio de mi propia experiencia consciente mientras estaba sentado en mi mecedora, mirando por la ventana en un hermoso día de primavera. Volvamos sobre ese pasaje y comprobemos cómo lo trata la teoría que he desarrollado. Éste es el texto:

> La luz verde-dorada de los primeros días de primavera penetraba a través de la ventana; las ramas y ramitas del arce del jardín eran todavía visibles a través de una nube de verdes brotes, formando un elegante dibujo de maravillosa complejidad. La vidriera de la ventana es de cristal antiguo y tiene un pequeño defecto que es apenas perceptible; mientras me balanceaba, esta imperfección provocó una onda de movimientos sincronizados que iba y venía a través del delta de ramas, un movimiento regular que se superponía con gran viveza sobre el desordenado y trémulo reflejo de las ramas movidas por la brisa.
>
> Entonces me di cuenta de que este metrónomo visual en las ramas del árbol seguía acompasadamente las notas del *Concerto grosso* de Vivaldi que había escogido como «música de ambiente» para mi lectura. (...) Mi pensar consciente y, especialmente, el deleite que me producía esa combinación de luz solar,

* *El mercader de Venecia*, de William Shakespeare, Acto III, Escena 1. Traducción castellana de Manuel Ángel Conejero, Juan Vicente Martínez Luciano y Jenaro Talens, Cátedra (Letras Universales 9), Madrid, 1984. [N. del T.]

6. ¡Citar estas palabras fuera de contexto sería un acto de desesperada falta de honestidad intelectual!

> soleados violines de Vivaldi y ramas susurrantes, además del placer que sentía al pensar en todo ello, ¿cómo es posible que *todo esto* no fuera más que algo físico que estaba ocurriendo en mi cerebro? ¿Cómo es posible que una determinada combinación de sucesos electroquímicos en mi cerebro pudiera sumarse a la delicadeza con que cientos de ramitas se balanceaban al ritmo de la música? ¿Cómo es posible que un acto de procesamiento de la información en mi cerebro *fuese* la suave calidez con que yo sentía la luz del sol sobre mi cuerpo? (...) Parece realmente imposible.

Habida cuenta de que he estado animando a que todos seamos heterofenomenólogos, yo no puedo quedar exento, y debo aceptar ser tanto sujeto como practicante, así que, ahí va, voy a aplicar mi propia teoría sobre mí mismo. Como heterofenomenólogos, nuestra tarea es tomar este texto, interpretarlo, y después relacionar los objetos del mundo heterofenomenológico de Dennett resultante con los acontecimientos que se están produciendo en el cerebro de Dennett en ese momento.

Dado que el texto se produjo unas semanas o meses después de que ocurrieran los acontecimientos de los que hablo, podemos estar seguros de que está resumido, no sólo a causa de la necesidad de condensación que los requisitos editoriales impusieron al autor, sino también por los inexorables procesos de resumen que la memoria impone en el tiempo. Si hubiéramos sondeado antes —si el autor hubiese tenido una grabadora a mano cuando estaba sentado en la mecedora, y hubiese producido el texto en ese lugar y en ese momento—, el texto sería, sin lugar a dudas, muy diferente. No sólo más rico en detalles, y más caótico, sino también reformado y redirigido por las propias reacciones del autor ante el proceso mismo de ir creando el texto: escuchando los sonidos de sus propias palabras en vez de reflexionar en silencio. Como sabe todo conferenciante, hablar en voz alta a menudo revela implicaciones (y problemas) en el propio mensaje que se escapan cuando uno se enfrasca en un soliloquio silencioso.

Tal como está, el texto retrata sólo una porción (una porción idealizada, sin duda) del contenido de la conciencia del autor. Debemos ser muy cuidadosos, sin embargo, de no suponer que «las partes que faltan» en el texto «estaban presentes» en algo que podríamos denominar flujo de la conciencia del autor. No debemos cometer el error de suponer que existen algunos hechos —hechos irrecuperables aunque reales— cuyos contenidos fueron conscientes en aquel momento y otros cuyos contenidos no lo fueron. En particular, no debemos suponer que cuando él miró por la ventana, «se lo tragó todo» de un solo y maravilloso bocado mental, aunque sea eso lo que el texto describe. De acuerdo con el texto, le pareció como si su mente —su campo visual— se llenara de complejos detalles de brotes verde-dorados y de un ir y venir de ramas, pero aunque eso es lo que le pareció, no fue más que una ilusión. Nunca se produjo un «pleno» así en su mente; el pleno se produjo fuera, en el mundo, allí donde no tenía que *representarse*, pues le bastaba con *ser*. Cuando nos maravillamos, en estos momentos de intensificada autoconciencia, de la gloriosa riqueza de nuestra experiencia consciente, la

riqueza que nos maravilla no es otra cosa que la riqueza del mundo exterior, con todo *su* detallismo embelesador. No «entra» en nuestras mentes conscientes, simplemente está disponible.

¿Qué podemos decir de las ramas y ramitas ondulándose al unísono? Las ramas del árbol de ahí afuera seguro que no se ondulaban así, ya que ese efecto se debía al defecto de la vidriera, lo cual tampoco significa que ese ondular se produjera en la mente o el cerebro del autor: se produjo en el interior de la vidriera que lo causó. Si alguien hubiese filmado las imágenes cambiantes en las retinas del autor, habría encontrado ese efecto en ellas, como en una película, pero no cabe duda de que el ondular acababa ahí; lo que ocurría en el interior de sus retinas no era más que el reconocimiento *de que* había, como él mismo dice en el texto, una maravillosa onda de movimientos que él podía experimentar. Él vio las ondas, y vio hasta dónde se extendían, del mismo modo que usted vería todas las Marilyns en el papel pintado. Y ya que sus retinas recibieron una dosis uniforme de ondulación, si le hubiese apetecido prolongar la experiencia *habría habido* más detalles en las Versiones Múltiples de las cuales sólo nos queda el texto que tenemos.

Hubo muchos más detalles en los que el autor podría haberse fijado, pero no lo hizo. Existen numerosos hechos decisivos, irrecuperables pero genuinos, sobre cuáles de estos detalles fueron discriminados, en qué momento y en qué lugar por diversos sistemas en el cerebro, pero la suma de todos estos hechos no nos permite responder de forma definitiva a la pregunta: ¿de qué elementos fue definitiva y realmente consciente (pero ya había olvidado en el momento de producir el texto), y cuáles permanecieron, definitiva y realmente, en un «segundo plano» de su conciencia (aunque no les prestó atención en ese momento)? Nuestra tendencia a suponer que *debe* haber un hecho decisivo para determinar estas cuestiones equivale a la suposición del lector inocente de que *debe* haber una respuesta para preguntas como: ¿había desayunado Sherlock Holmes huevos con jamón el día en que el doctor Watson lo conoció? Conan Doyle *podría* haber incluido este detalle en el texto, pero no lo hizo, y, puesto que no lo hizo, no existe ningún hecho decisivo que nos permita determinar si esos huevos con jamón pertenecen o no al *mundo ficticio de Sherlock Holmes*. Aun en el caso de que Conan Doyle hubiera pensado en Holmes comiendo huevos con jamón aquella mañana, aun en el caso de que en una primera versión del texto *se representara en forma de palabras manuscritas* a Holmes comiendo huevos con jamón aquella mañana, no existe ningún hecho decisivo sobre si en el mundo ficticio de Sherlock Holmes, el mundo constituido a partir del texto publicado que ha llegado hasta nosotros, éste desayunó huevos con jamón.

El texto de Dennett de que disponemos no permaneció «escrito en su cerebro» en el tiempo que media entre su experiencia en la mecedora y el momento en que lo escribió en un archivo con su procesador de textos. El estado de atención en que entró mientras se mecía, y la repetición concomitante de aquellos detalles particulares que más llamaron su atención, tuvieron el efecto de fijar, con relativa seguridad, el contenido de esos detalles particu-

lares «en memoria», aunque este efecto no debería interpretarse como el almacenamiento de una imagen (o una oración) o cualquier otra representación destacable. Por el contrario, sólo debemos pensar en ello como el acto de aumentar las posibilidades de que una actividad similar se reprodujese, algo que podemos suponer que finalmente ocurrió en el momento de redactar el texo: se reprodujo esa actividad y ello tuvo como consecuencia que una serie de demonios-palabra formaran coaliciones que, por primera vez, dieron lugar a una cadena de palabras. No cabe duda, sin embargo, de que parte de lo que ocurrió antes, en la mecedora, activó algunas palabras y frases inglesas, ni de que esa colaboración inicial entre contenidos sin palabras y palabras facilitó, en el momento de redactar el texto, la recuperación de algunas de esas mismas expresiones inglesas.

Volvamos al mundo heterofenomenológico del texto. ¿Qué podemos decir del gozo de que nos habla? «...esa combinación de luz solar, soleados violines de Vivaldi y ramas susurrantes, además del placer que sentía al pensar en todo ello...» Conceptos como éstos no podrían explicarse apelando a qualia de visión, de sonido y de puro pensamiento intrínsecamente placenteros. La idea de que dichos qualia existen no hace más que distraernos de todas las demás vías posibles de explicación, captando nuestra atención de la misma manera que el movimiento de un dedo capta la atención de un bebé, haciendo que nos quedemos embobados contemplando ese «objeto intrínseco» en vez de andar buscando una descripción de los mecanismos subyacentes y una explicación (en última instancia una explicación en términos evolutivos) de por qué esos mecanismos hacen lo que hacen.

El gozo del autor es fácil de explicar por el hecho de que toda experiencia visual se compone de actividades de circuitos neuronales cuya actividad misma nos resulta innatamente placentera, no sólo porque nos gusta ser informados, sino también porque nos gustan las diferentes maneras de quedar informados. No debería sorprendernos que la visión de unos brotes primaverales bañados por la luz del sol resulte agradable para un humano. El que a ciertos humanos también les guste mirar imágenes microscópicas de las bacterias o fotografías de catástrofes aéreas es mucho más raro, pero las sublimaciones y las perversiones del deseo surgen de las mismas fuentes animales, la configuración de nuestros sistemas nerviosos.

El autor prosigue preguntándose «cómo es posible que todo eso no sea más que la combinación de unos acontecimientos electroquímicos en mi cerebro». Como demuestra su pregunta, no parece que sea así. O, en todo caso, hubo un momento en que tuvo la sensación de que no le parecía ser una mera combinación de acontecimientos electroquímicos en su cerebro. Sin embargo, lo dicho en capítulos subsiguientes sugiere un réplica: bueno, ¿pues qué piensa usted que parecería si *fuera* una mera combinación de acontecimientos electroquímicos en su cerebro?[7] ¿Acaso no hemos sentado las bases para

7. Véase Lockwood (1989): «¿Cómo se percibiría la conciencia si se *percibiera* como miles de millones de átomos agitándose en su interior?» (págs. 15-16)

llegar a la conclusión de que con un cerebro que esté organizado como el nuestro, éste es precisamente el tipo de mundo heterofenomenológico que debemos esperar? ¿Por qué motivo las combinaciones de acontecimientos electroquímicos en el cerebro no podrían tener precisamente los efectos que nos dispusimos a explicar?

(Habla el autor): queda todavía un enigma, sin embargo. ¿De qué modo llego *yo* a saber todo esto? ¿Cómo es posible que *yo* pueda contar lo que está ocurriendo en mi cabeza? La respuesta a este enigma es fácil: *porque esto es lo que yo soy*. Porque el que conoce y refiere estas cosas de esta manera es lo que yo soy. *Mi* existencia se explica por el hecho de que en este cuerpo existen estas capacidades.

Esta idea, la idea del yo como centro de gravedad narrativa, es una idea que ya estamos en disposición de tratar. Ya no cabe duda de que a esta idea le ha llegado su momento. Imaginen mis sentimientos encontrados cuando descubrí que antes de que yo consiguiera ver mi propia versión de la misma debidamente publicada en un libro,[8] ya había sido satirizada en una novela, *Nice Work* de David Lodge (1988). Aparentemente, es un asunto que motiva calurosos debates entre los desconstruccionistas:

> Según Robyn (o, más concretamente, según los autores que han influido en su pensamiento sobre estas cuestiones), no existe ese «yo» sobre el que el capitalismo y la novela clásica se han basado; es decir, no existe esa alma o esencia única y finita que constituye la identidad de la persona. Sólo existe la posición de un sujeto en un tejido infinito de discursos: los discursos del poder, del sexo, de la familia, de la ciencia, de la religión, de la poesía, etc. Y por el mismo motivo, tampoco existe eso que llamamos autor, es decir, ese alguien que origina una obra de ficción *ab nihilo*. (...) parafraseando las famosas palabras de Jacques Derrida... «*il n'y a pas de hors-texte*», no hay nada fuera del texto. No hay orígenes, sólo hay producción, y nosotros producimos nuestros «yos» en el lenguaje. No «*tú eres lo que tú comes*», sino «*tú eres lo que tú hablas*», o, mejor, «*tú eres lo que te habla*», ésa es la base axiomática de la filosofía de Robyn, que ella denominaría, si es que tuviera que ponerle un nombre, «materialismo semiótico».

¿Materialismo semiótico? ¿Es así como tengo que llamarlo *yo*? Aparte de las alusiones al capitalismo y a la novela clásica, sobre los cuales yo no me he pronunciado, este divertido pasaje es una excelente parodia de la visión que estoy a punto de presentar. (Como todas las parodias, ésta también

8. Presenté mis ideas básicas sobre el asunto en mis reflexiones sobre Borges, en *The Mind's I* (Hofstadter y Dennett, 1981, págs. 348-352), y luego las reuní en el texto de una conferencia, «The Self as the Center of Narrative Gravity», pronunciada en el Houston Symposium en 1983. Mientras esperaba que se publicaran las actas de este simposio, publiqué una versión resumida de mi conferencia en el *Times Literary Supplement*, 16-22 de septiembre, 1988, bajo el aburridísimo título —que yo no escogí— de «Why everyone is a novelist». La versión original, titulada «The Self as the Center of Narrative Gravity» sigue sin haber aparecido hasta la fecha en F. Kessel, P. Cole y D. Johnson, comps., *Self and Consciousness: Multiple Perspectives*, Erlbaum, Hillsdale, NJ. [El texto apareció, finalmente, en 1992; N. del T.]

exagera; yo no diría que no hay *nada* fuera de texto. Por ejemplo, fuera de texto están las librerías, los edificios, los cuerpos, las bacterias...)

Robyn y yo pensamos de manera parecida, y, evidentemente, *ambos* somos, de acuerdo con nuestras propias visiones, una especie de personajes ficticios, aunque de especies ligeramente distintas.

Capítulo 13

LA REALIDAD DE LOS YOS

> Y si imaginamos una máquina cuya estructura haga pensar, sentir, tener percepción, se la puede concebir de mayor tamaño conservando las mismas proporciones, de manera que se pueda entrar en ella como en un molino. Concedido esto, al visitarla por dentro sólo se hallarían piezas que se empujan unas a otras y jamás algo con que explicar una percepción.
>
> GOTTFRIED WILHELM LEIBNIZ (1646-1716), *Monadología* (publicado por primera vez en 1840)*

> En lo que a mí respecta, siempre que penetro más íntimamente en lo que llamo *mí mismo* tropiezo en todo momento con una u otra percepción particular, sea de calor o frío, de luz o de sombra, de amor u odio, de dolor o placer. Nunca puedo atraparme a *mí mismo* en ningún caso sin una percepción, y nunca puedo observar otra cosa que la percepción. (...) Si tras una reflexión seria y libre de prejuicios hay alguien que piense que él tiene una noción diferente de *sí mismo*, tengo que confesar que yo no puedo seguirle en sus razonamientos. Todo lo que puedo concederle es que él puede estar tan en su derecho como yo, y que ambos somos esencialmente diferentes en este particular. Es posible que él pueda percibir algo simple y continuo a lo que llama su *yo*, pero yo sé con certeza que en mí no existe tal principio.
>
> DAVID HUME, *Tratado de la naturaleza humana*, 1739.*

Desde los albores de la ciencia moderna en el siglo XVIII, ha existido un acuerdo unánime sobre el yo: sea lo que sea, éste sería invisible al microscopio y también invisible a la introspección. Para algunos, esto ha sido motivo suficiente para suponer que el yo era un alma no física, un espíritu en la máquina. Otros, se han visto inclinados a pensar que el yo no era nada, una quimera producto de imaginaciones metafísicamente febriles. Y finalmente, otros han pensado que el yo era, de un modo u otro, una especie de abstracción, algo cuya existencia no se veía afectada en lo más mínimo por su invisibilidad. Después de todo, podemos argumentar que un centro de

* *Monadología*, §17. Traducción de E. de Olaso, en *Escritos Filosóficos*, Ed. Charcas, Buenos Aires, 1982, pág. 610. [N. del T.]

* *Tratado de la naturaleza humana*, Libro I: Del entendimiento, Parte IV: Del escepticismo y otros sistemas de filosofía, Sección VI: De la identidad personal. Traducción y edición de Félix Duque, Editora Nacional, Madrid, 1980-1981 (2 vols.), págs. 399-400 y (revisada), Tecnos, Madrid, 1988, págs. 355-356. [N. del T.]

gravedad es igualmente invisible, e igualmente real. ¿Es esto suficientemente real?

Podemos hacer en cualquier caso que la pregunta de si realmente existen los yos parezca algo ridículo a lo que responder: ¿Existimos *nosotros*? ¡Claro que sí! La pregunta presupone la respuesta. (Después de todo, ¿quién es ese *yo* que, según Hume, buscaba en vano un yo?) ¿Existen entidades, *en* nuestros cerebros, o *además* de nuestros cerebros, que controlen nuestros cuerpos, piensen nuestros pensamientos, tomen nuestras decisiones? ¡Claro que no! Esta idea es o una idiotez empírica (la «neurona pontificia» de James) o una estupidez metafísica (el «espíritu en la máquina» de Ryle). Cuando ante una pregunta tan simple se pueden dar dos respuestas, «¡claro que sí!» y «¡claro que no!», suele ser recomendable considerar una posición intermedia (Dennett, 1991a), aunque esté condenada a resultar profundamente cotraintuitiva para los partidarios de una u otra respuesta; ¡todos coincidirán en que niega algún hecho evidente u otro!

1. Cómo tejen los humanos un yo

> Además, parecían pasar la mayor parte de su tiempo comiendo y bebiendo y asistiendo a fiestas, y Frensic, cuya apariencia física tendía a limitar sus placeres sensuales a meterse cosas a sí mismo en vez de metérselas a los demás, era un verdadero *gourmet*.
>
> TOM SHARPE, *The Great Pursuit*, 1977

El novelista Tom Sharpe nos sugiere, en este divertido pero turbador pasaje, que, cuando llegamos al fondo del asunto, todo placer sensual no consiste más que en jugar con nuestros propios límites o con los límites de los demás; y parece tener razón: si no toda la razón, al menos sí parte de razón.

Las personas tienen yos. ¿Los tienen los perros? ¿Y las langostas? Si los yos son realmente algo, entonces existen. *Ahora*, los yos existen. Hubo un tiempo, hace miles (o millones o miles de millones) de años, en que los yos no existían, cuando menos no existían en este planeta. Así pues, por lógica, debe haber una historia que podamos contar sobre *cómo llegó a haber* criaturas con un yo. Esta historia nos tendrá que hablar, por lógica, de un proceso (o una serie de procesos) en el que participasen las actividades y los comportamientos de cosas que todavía no *poseían* un yo —o que todavía no *eran* un yo—, pero que, finalmente, dio lugar, como producto nuevo, a seres que son, o poseen, un yo.

En el capítulo 7 vimos que el nacimiento de las razones también fue el nacimiento de los límites, el límite entre «yo» y «el resto del mundo», una distinción que debe hacer hasta la más simple ameba, a su manera ciega y falta de conocimiento. Esta mínima inclinación a distinguir el yo del otro a fin de protegerse a uno *mismo* es el yo biológico, e incluso este yo tan simple no es una cosa concreta, sino una abstracción, un principio de organiza-

ción. Además, los límites del yo biológico son porosos e indefinidos, otro ejemplo de la tolerancia al «error» de la Madre Naturaleza, si el coste lo justifica.

Dentro de las murallas de los cuerpos humanos habitan muchos, muchos intrusos, desde bacterias y virus hasta parásitos mayores —como, por ejemplo, las horribles tenias—, pasando por los microscópicos ácaros que viven, como los habitantes de un bloque de pisos, en el nicho ecológico de nuestra piel y nuestro cuero cabelludo. Estos intrusos trabajan todos por la protección de sus propios derechos, pero algunos de ellos, como las bacterias que pueblan nuestro sistema digestivo y sin las cuales moriríamos, son unos miembros del equipo tan imprescindibles para nuestra propia conservación como los anticuerpos de nuestros sistemas inmunológicos. (Si la teoría de la bióloga Lynn Margulis [1970] es correcta, las mitocondrias que hacen el trabajo en casi todas las células de nuestro cuerpo serían las descendientes de unas bacterias con las cuales «nosotros» unimos nuestras fuerzas hace unos dos mil millones de años.) Otros intrusos son parásitos tolerados —que, aparentemente, no vale la pena desahuciar—, mientras que otros son el verdadero enemigo interior, mortales si no se los expulsa.

Este principio biológico fundamental consistente en distinguir al yo del mundo, al interior del interior, resuena con un fuerte eco en las cúpulas más altas de nuestra psicología. Los psicólogos Paul Rozin y April Fallon (1987), en una serie de fascinantes experimentos sobre la naturaleza del *asco*, han demostrado que existe una poderosa y desconocida corriente subterránea de resistencia ciega a llevar a cabo ciertos actos que, considerados desde un punto de vista racional, no deberían provocarnos ningún rechazo. Por ejemplo, ¿podría usted tragarse la saliva que tiene en su boca, por favor? Este acto no nos provoca ninguna repulsión. Pero suponga que yo le hubiera pedido que tomara un vaso limpio, que escupiera en él y que, *después*, se tragara la saliva del vaso. ¡Repugnante! ¿Pero, por qué? Parece tener que ver con nuestra percepción de que una vez algo ha abandonado nuestro cuerpo, ya no forma parte de nosotros —se convierte en algo ajeno y sospechoso—, ha renunciado a su ciudadanía y se convierte en algo que no merece más que rechazo.

La superación de los límites coincide, por tanto, con momentos de profunda ansiedad, o, como señala Sharpe, con momentos de enorme placer. Muchas especies han desarrollado unas construcciones notables para ampliar los límites territoriales, sea para dificultar las transgresiones negativas sea para facilitar las transgresiones positivas. Los castores construyen presas, y las arañas tejen telas, por ejemplo. Cuando una araña teje su tela, no tiene por qué comprender lo que está haciendo; la Madre Naturaleza se ha limitado a equipar su minúsculo cerebro con las rutinas necesarias para ejecutar esta tarea de ingeniería tan fundamental desde el punto de vista biológico. Los experimentos con castores demuestran que incluso sus magníficas y eficientes capacidades como ingenieros son, en gran medida, el producto de impulsos e inclinaciones innatos que esos animales no necesitan comprender para beneficiarse de ellos. Los castores aprenden, e incluso pue-

den instruirse los unos a los otros, pero se rigen principalmente por unos poderosos mecanismos que controlan lo que el conductista B. F. Skinner denominó refuerzo negativo. Un castor buscará, frenéticamente incluso, algo —cualquier cosa— que le sirva para detener el sonido del agua corriente: en un experimento, un castor pudo aliviar sus impulsos cubriendo por completo de barro ¡el altavoz de donde salía el sonido grabado del gorgoteo de agua! (Wilsson, 1974)

El castor proteje sus límites exteriores con ramitas y barro, y uno de sus límites interiores con su piel. El caracol acumula el calcio de su alimento y lo utiliza para producir una concha dura; el cangrejo ermitaño utiliza una concha de calcio prefabricada, apropiándose de la caracola de otra criatura y ahorrándose, así, el proceso de ingestión y producción. La diferencia no es fundamental, según Richard Dawkins, quien señala que, en todos los casos, el resultado, que él denomina el *fenotipo ampliado* (1982), forma parte del equipo biológico fundamental de los individuos sometidos a las fuerzas selectivas que gobiernan la evolución.

La definición de fenotipo ampliado no sólo va más allá de los límites «naturales» de los individuos, lo que incluye equipos externos tales como las conchas (y equipos internos tales como las bacterias residentes); sino que a veces incluye también a otros individuos de la misma especie. Los castores no pueden actuar solos, necesitan trabajar en equipo para construir una presa. Las termitas tienen que agruparse por millones para construir sus castillos.

Y considérense las sorprendentes obras arquitectónicas del pájaro jardinero australiano *(Ptilonorhynchidae)* (Borgia, 1986). Los machos construyen unas complejas glorietas, santuarios del cortejo con grandes naves centrales, ricamente decoradas con objetos de colores brillantes, predominantemente azules, como chapas de botella, trocitos de cristal coloreado y muchos otros objetos de manufactura humana, que recogen en lugares lejanos y disponen cuidadosamente en su glorieta para impresionar a la hembra que están cortejando. El jardinero, como la araña, tampoco tiene que comprender lo que está haciendo; simplemente trabaja duramente, sin saber por qué, para construir un edificio que es crucial para su éxito como pájaro-glorieta.

Pero las construcciones más extrañas y maravillosas de todo el mundo animal son las increíbles y complejas construcciones que levanta un primate, el *Homo sapiens*. Todo individuo normal de esta especie construye un yo. A partir de su cerebro teje una tela de palabras y de actos, y, como las demás criaturas, no tiene por qué saber qué está haciendo; sólo lo hace. Esa tela lo proteje, como la concha del caracol, le proporciona el sustento, como la tela de la araña, y favorece sus perspectivas para el sexo, como la glorieta del pájaro jardinero. Al contrario que la araña, sin embargo, un ser humano no *exuda* su tela, sino que, como un castor, trabaja muy duro para recoger los materiales con los que construir su fortaleza. Como un pájaro jardinero, se apropia de aquellos objetos que le gustan —a él o a su pareja—, incluidos muchos que fueron diseñados por otros con propósitos muy distintos.

Ese «tejido de discursos», como Robyn lo bautizó al final del capítulo anterior, es tanto un producto biológico como lo son las demás construcciones que encontramos en el mundo animal. Separado de él, un ser humano está tan incompleto como un pájaro sin sus plumas o como una tortuga sin su concha. (La vestimenta también es parte del fenotipo ampliado del *Homo sapiens* en casi todos los nichos habitados por esta especie. Representar desnudo al *Homo sapiens* en una enciclopedia ilustrada de zoología tiene el mismo sentido que representar al *Ursus arctus* —el oso negro— montado en bicicleta y vestido de payaso.)

Tan maravillosa es la organización de una colonia de termitas, que a muchos observadores les pareció que toda colonia de termitas tenía que tener un alma (Marais, 1937). Ahora sabemos que esa organización se debe simplemente a la presencia de un millón de agentes semiindependientes, cada uno un autómata por sí mismo, que llevan a cabo sus tareas. Tan maravillosa es la organización de un yo humano, que a muchos observadores también les ha parecido que todo ser humano tiene que tener un alma: un dictador benevolente gobernando desde su cuartel general.

En todo panal de abejas o colonia de termitas hay una abeja reina o una termita reina, pero estos individuos son muchos más pasivos que un agente, son más como las joyas de la Corona que hay que proteger que como el jefe de las fuerzas de defensa; de hecho su carácter real encaja mucho mejor con la actualidad que con el pasado, ya que se parecen mucho más a la reina Isabel II que a la reina Isabel I. No hay ninguna abeja Margaret Thatcher, ni ninguna termita George Bush, ni ningún Despacho Oval en el hormiguero.

¿Muestran nuestros yos, nuestros máximamente *egoicos* yos, la misma permeabilidad y flexibilidad en cuanto a los límites que los yos más simples de otras criaturas? ¿Ampliamos nuestros límites personales —los límites de nuestros *yos*— a fin de abarcar parte de nuestra «sustancia»? En general quizá no, pero desde el punto de vista psicológico hay momentos en que esto parece ser verdad. Por ejemplo, hay personas que se limitan a tener un coche y conducirlo; otras, en cambio, son *automovilistas*; el automovilista empedernido prefiere *ser* un agente con cuatro ruedas y consumidor de gasolina que un agente con dos piernas y consumidor de alimentos, y su uso de la primera persona suele reflejar esta identificación:

> «Últimamente no tomo bien las curvas cuando llueve porque tengo los neumáticos muy gastados.»

Así que, ocasionalmente, ampliamos nuestros límites; en otras ocasiones, en cambio, como respuesta a ciertos desafíos, reales o imaginarios, reducimos nuestros límites:

> «¡*Yo* no hago esto! No era mi verdadero yo el que hablaba. Sí, sí, las palabras surgieron de mi boca, pero no puedo reconocerlas como mías.»

He traído a colación estos discursos a fin de remarcar el parecido existente entre nuestros yos y los yos de las hormigas y los cangrejos ermitaños, pero dichos discursos también me permiten llamar la atención sobre una diferencia fundamental: las hormigas y los cangrejos ermitaños no hablan. El cangrejo ermitaño ha sido diseñado a fin de asegurar que se procure una una concha. Su organización, podríamos decir, *implica* la concha, y, por tanto, en un sentido *muy* amplio, representa tácitamente al cangrejo como poseedor de una concha, pero el cangrejo, en ningún sentido estricto del término, se representa a sí mismo como poseedor de una concha. De ningún modo es un candidato para la autorrepresentación. ¿Para quién *se representaría a sí mismo* y por qué? No necesita recordarse a sí mismo este aspecto de su naturaleza, ya que su diseño innato se ocupa del asunto, y no hay nadie más en los alrededores que esté interesado. Y las hormigas y las termitas, como hemos visto, llevan a cabo sus proyectos comunitarios sin basarse en ningún anteproyecto o edicto que se les haya comunicado explícitamente.

Nosotros, por el contrario, estamos casi constantemente ocupados en presentarnos a nosotros mismos a los demás, o a nosotros, y, por tanto, en *representarnos* a nosotros mismos, con el lenguaje y el gesto, externo e interno. La diferencia más evidente dentro de nuestro entorno que explicaría esta diferencia de conducta es la conducta misma. Nuestro entorno humano no contiene solamente alimento y cobijo, enemigos con los que luchar y de los que escapar, y miembros de la misma especie con los que aparearnos, sino palabras, palabras, palabras. Estas palabras son unos poderosos elementos de nuestro entorno que incorporamos fácilmente, ingiriéndolas y excretándolas, tejiéndolas como telas de araña hasta construir secuencias de *narraciones* autoprotectoras. Evidentemente, como vimos en el capítulo 7, cuando permitimos la entrada a estas palabras, a estos vehículos para memas, éstas toman el mando, creándo*nos* a partir de las materias primas que encuentran en nuestros cerebros.

Nuestra táctica fundamental de autoprotección, de autocontrol y de autodefinición no consiste en tejer una tela o construir una presa, sino en contar historias, y más particularmente, en urdir y controlar la historia que contamos a los demás —y a nosotros mismos— sobre quiénes somos. Y al igual que las arañas no tienen que pensar, consciente y deliberadamente, en cómo deben tejer sus telas, y al igual que los castores, a diferencia de los *profesionales* humanos de la ingeniería, no proyectan consciente y deliberadamente las estructuras que construyen, nosotros (a diferencia de los *profesionales* humanos del contar historias) no imaginamos consciente y deliberadamente qué narraciones contar ni cómo contarlas. Nuestras historias se urden, pero en gran parte no somos nosotros quienes las urdimos; ellas nos urden a nosotros. Nuestra conciencia humana, nuestra egoticidad narrativa, es su producto, no su origen.

Estas secuencias o flujos narrativos surgen *como si* fueran emitidos por una misma fuente, no en el claro sentido físico de surgir de una boca, de un lápiz o de una pluma, sino en un sentido más sutil: su efecto sobre una

audiencia es el de animarla a (intentar) postular un agente unificado a quien pertenecen esas palabras y sobre quien son esas palabras: es decir, la animan a postular un *centro de gravedad narrativa*. Los físicos aprecian la enorme simplificación que se obtiene al postular el centro de gravedad de un objeto, un único punto en relación al cual todas las fuerzas gravitatorias pueden ser calculadas. Nosotros, los heterofenomenólogos, apreciamos la enorme simplificación que se obtiene al postular un centro de gravedad narrativa para el tejido narrativo de un cuerpo humano. Como el yo biológico, este yo narrativo o psicológico es otra abstracción, no una cosa en el cerebro, pero, con todo, es un atraedor de propiedades muy robusto y casi tangible, el «propietario del registro» de todos aquellos elementos y aquellos rasgos que no han sido reclamados. ¿Quién es el dueño de su coche? Usted. ¿Quién es el dueño de su ropa? Usted. Entonces, ¿quién es el dueño de su cuerpo? ¡Usted! Cuando usted dice,

«Éste es *mi* cuerpo»,

seguro que usted no está diciendo,

«este cuerpo se posee a sí mismo».

¿Pero qué está usted diciendo entonces? Si lo que usted está diciendo no es una extravagante tautología carente de sentido (este cuerpo es su propio propietario, o algo por el estilo) ni la afirmación de que usted es un alma inmaterial o un titiritero espiritual que posee y hace funcionar este cuerpo del mismo modo que usted posee y hace funcionar su coche, ¿qué otra cosa puede querer decir?

2. ¿Cuántos yos por cliente?

Creo que podríamos ver con más claridad lo que significa

«Éste es mi cuerpo»,

si pudiéramos responder a la siguiente pregunta: ¿en oposición a qué? ¿Qué me dicen de esto?

«No, no lo es; es *mío*, ¡y no me gusta compartirlo!»

Si pudiéramos ver cómo sería que dos (o más) yos compitieran por el control de un único cuerpo, podríamos comprender mejor lo que es realmente un único yo. Como investigadores del yo, nos gustaría llevar a cabo experimentos controlados, en los que, variando las condiciones iniciales, pudiéramos observar lo que tiene que ocurrir, en qué orden y con qué requisitos,

para que emerja un yo parlante. ¿Existen condiciones bajo las cuales la vida continúa sin que llegue a emerger un yo? ¿Existen condiciones bajo las cuales puede emerger más de un yo? Éticamente no podemos llevar a cabo estos experimentos, pero, como muchas otras veces, podemos aprovechar los datos generados por algunos de esos terribles experimentos que hace la naturaleza y llegar, con mucha cautela, a algunas conclusiones.

Los Trastornos de Personalidad Múltiple (TPM) serían ese experimento, en que un único cuerpo humano *parece* ser compartido por más de un yo, cada uno con su nombre propio y su autobiografía. La idea del TPM se le antoja a muchas personas como algo demasiado estrafalario y metafísicamente extravagante como para ser creíble, un fenómeno «paranormal» equiparable a la PES (Percepción Extrasensorial), los encuentros en la tercera fase y las brujas montadas en una escoba. Sospecho que algunas de estas personas han cometido un simple error aritmético: no han sabido darse cuenta de que dos, tres o diecisiete yos por cuerpo no es más metafísicamente extragavante que un yo por cuerpo. ¡Ya es bastante grave que haya uno!

> «Acabo de ver un coche con cinco yos dentro.»
> «¿¿Qué?? ¡La cabeza me da vueltas! ¡Menudo sinsentido metafísico!»
> «Bueno, la verdad es que en el coche también había cinco cuerpos.»
> «Haber empezado por ahí, ¿por qué no lo has dicho antes? Ahora todo está claro.»
> «O, quizá sólo había cuatro cuerpos, o tres... en cualquier caso seguro que había cinco yos.»
> «¡¡¿¿Qué??!!»

Lo normal es un yo por cuerpo, pero si un cuerpo puede tener un yo, ¿por qué no podría tener más de un yo en condiciones anormales?

No pretendo decir con ello que no haya nada chocante o profundamente enigmático en los TPM. Es, de hecho, un fenómeno de una gran extrañeza, pero no, creo, porque constituya un desafío a nuestras presuposiciones sobre lo que es *metafísicamente* posible, sino porque es un desafío a lo que nos parece *humanamente* posible, en cuanto a los límites de la crueldad y la depravación humana, por un lado, y en cuanto a los límites de la creatividad humana, por el otro. Pues no podemos más que rendirnos ante la evidencia de que los TPM diagnosticados hasta hoy no son unos pocos cientos, sino muchos miles, los cuales son atribuibles, casi invariablemente, a abusos sufridos durante la infancia, generalmente de tipo sexual, de una gravedad espantosa, hace algunos años. Nicholas Humphrey y yo mismo estudiamos los TPM (Humphrey y Dennett, 1989) y descubrimos que se trata de un fenómeno muy complejo que se extiende mucho más allá de los cerebros de aquellos individuos que los padecen.

Estos niños han vivido a veces unas circunstancias tan terribles y confusas que me sorprende más el hecho de que, psicológicamente, consigan sobrevivir, de lo que me sorprende que consigan conservarse mediante una de-

sesperada reconstrucción de sus límites. Cuando tienen que enfrentarse a un dolor y a un conflicto abrumadores, lo que hacen es esto: «se marchan». Crean un límite de tal modo que el horror no les afecta a *ellos*: o no afecta a nadie o afecta a otro yo, más capacitado para soportar esa violencia; o, cuando menos, eso es lo que *dicen* que hicieron, cuando lo recuerdan.

¿Cómo es esto posible? ¿Qué explicación —que, en última instancia, debería ser de tipo biológico— podemos dar de este proceso de separación? ¿Se trata acaso de un único yo completo, que se divide, como si fuese una ameba? ¿Cómo es esto posible, si un yo no es parte física de un organismo o de un cerebro, sino, como he propuesto, una abstracción? La respuesta al trauma es tan creativa, asimismo, que uno se siente inclinado a creer en un principio que detrás de todo está la mano de una especie de supervisor: un programa cerebral de supervisión, un controlador central, o algo por el estilo. No debemos olvidar, no obstante, la colonia de termitas, que también parecía necesitar, en un primer momento, un mando ejecutivo central para llevar a cabo tan inteligentes proyectos.

Nos hemos acostumbrado a oír narraciones evolucionistas que empiezan con un estadio en que un fenómeno determinado todavía no existe y acaban con un estadio en que el fenómeno ya está presente: innovaciones como la agricultura, el vestido, la vivienda y las herramientas, la innovación del lenguaje, la innovación de la conciencia misma o la innovación aún más temprana de la vida en la tierra. Todas estas historias están ahí para ser contadas. Y cada una de ellas tiene que atravesar lo que podríamos denominar la sima del absolutismo. Podemos ejemplificar este abismo con el curioso argumento siguiente (que tomo prestado de Sanford, 1975):

> Todo mamífero tiene un mamífero por madre,
> pero sólo hay un número finito de mamíferos, así que
> debe de haber existido un primer mamífero,
> lo cual contradice nuestra primera premisa, así que,
> y al contrario de lo que parece,
> ¡los mamíferos no existen!

Algo tiene que ceder. ¿Pero qué? El filósofo absolutista o esencialista se siente atraído por las líneas bien definidas, los umbrales, las «esencias» y los «criterios». Para el absolutista, siempre tiene que haber habido un primer mamífero, un primer ser viviente, un primer instante de la conciencia, un primer agente moral: el producto cualquiera de algún salto, un candidato cualquiera, por radical que sea, que cumpla las condiciones esenciales; cualquier cosa que revele el análisis.

Fue este gusto por los límites bien definidos entre las especies lo que constituyó uno de los mayores obstáculos intelectuales que Darwin tuvo que afrontar en su intento de desarrollar la teoría de la evolución (Richards, 1987). Opuesto a esta manera de pensar tenemos el tipo de antiesencialismo que se caracteriza por su gusto por los lugares en penumbra y la falta de líneas

divisorias estrictas. Dado que los yos y las mentes e incluso la conciencia son productos biológicos (y no elementos que podamos hallar en la tabla periódica de la química), debemos esperar que la transición entre estos fenómenos y los fenómenos que todavía no son ellos sea gradual, contenciosa, amañada. Ello no significa que todo esté siempre en transición, que siempre sea gradual; las transiciones que se perciben como graduales desde un punto de vista próximo, suelen verse como picos abruptos entre mesetas de equilibrio desde un punto de vista más distante (Eldredge y Gould, 1972; pero véase también Dawkins, 1982, págs. 101-109).

La importancia de este hecho para las teorías filosóficas (y las predilecciones de los filósofos) no ha recibido el reconocimiento que merece. Siempre ha habido —y siempre habrá— entidades de transición, «eslabones perdidos», cuasimamíferos y cosas parecidas que desafían todo intento de definición, pero el hecho es que *casi todas* las cosas reales (por oposición a lo que es meramente posible) de la naturaleza tienden a caer dentro de grupos de semejanza separados, en el espacio lógico, por gigantescos océanos de vacío. No necesitamos «esencias» ni «criterios» para asegurar que los significados de nuestras palabras no se deslicen de un lado para otro; nuestras palabras se estarán quietas, firmemente ancladas, como por gravedad, al grupo de semejanza más cercana, incluso si ha habido —debe de haber habido— un breve istmo que en algún momento las conectó, mediante una serie de pasos graduales, con alguna categoría vecina. Esta idea se aplica sin discusión a muchos otros asuntos. Pero muchas personas que con toda naturalidad aplican este enfoque pragmático al día y la noche, a lo viviente y lo no viviente, a los mamíferos y los premamíferos, se molestan cuando se les invita a aplicarlo también al caso de tener y no tener un yo. Piensan que aquí, al contrario que en todos los demás ámbitos de la naturaleza, se trata de «todo o nada» y «sólo uno por cliente».

La teoría de la conciencia que hemos venido desarrollando desacredita todos estos supuestos, y los Trastornos de Personalidad Múltiple constituyen un buen ejemplo de cómo la teoría los pone en cuestión. La convicción de que *no puede* haber cuasiyos o especies de yos, y de que, además, *debe* haber un número entero de yos asociado a un cuerpo —¡y mejor que ese número sea el uno!— *no* es evidente. Es decir, ya no es evidente, ahora que hemos desarrollado con cierto detalle una alternativa al Teatro Cartesiano con su testigo o Significador Central. Los TPM constituyen un desafío para estos supuestos por uno de los flancos, pero también podemos imaginar un desafío por el otro flanco: ¡dos o más cuerpos que compartan un único yo! Puede que también se haya producido un caso que se corresponda a esta posibilidad, en York, Inglaterra: las gemelas Chaplin, Greta y Freda (*Time*, 6 de abril de 1981). Estas gemelas idénticas, rozando actualmente la cuarentena, que viven juntas en un hotel, parecen actuar *como si hubiese una*; colaboran en la proferencia de actos de habla individuales, por ejemplo, terminando la una la frase que empezó la otra o hablando a la vez, con un desfase apenas perceptible. Durante años han sido inseparables, todo lo insepara-

bles que pueden ser dos gemelos que nos son siameses. Algunos de los que han tratado con ellas sugieren que la táctica más natural y efectiva que se puede adoptar es considerarlas como más de una *ella*.

Nuestra visión tolera la posibilidad teórica no sólo de los TPM, sino también de los TPF (Trastornos de Personalidad Fraccionada). ¿Puede ser? ¿Por qué no? No estoy sugiriendo, de momento, que las gemelas estén conectadas por telepatía o por PES o por cualquier otro tipo de vínculo oculto. Lo que estoy sugiriendo es que hay muchas maneras sutiles y cotidianas de comunicarse y de coordinarse (técnicas que suelen ser explotadas al máximo por los gemelos idénticos). Habida cuenta de que los gemelos han visto, oído, tocado, olido y pensado en casi los mismos acontecimientos a lo largo de sus vidas, que empezaron, sin duda, con cerebros dispuestos a reaccionar de forma muy similar a estos estímulos, podría ser que no se necesitaran unos canales de comunicación muy grandes para mantenerlos dentro de una especie de leve armonía. (Y además, ¿qué grado de unificación posee el más sereno de nosotros?) Deberíamos guardarnos de establecer los límites de esta capacidad de coordinación.

Pero, en cualquier caso, ¿no podría haber dos yos individuales perfectamente definidos, uno para cada gemela, y responsables de mantener esta curiosa charada? Quizá, pero, ¿y si cada una de estas mujeres hubiese llegado a estar tan desinteresada y carente de yo en su devoción por la causa común, que acabara por perder a su yo y a sí misma en el proyecto?* Como dijo una vez el poeta Paul Valéry, dando un nuevo giro a las palabras de su compatriota: «A veces soy, a veces pienso».

En el capítulo 11 vimos que mientras la conciencia parece ser continua, está, de hecho, llena de huecos. Un yo podría estar igualmente lleno de huecos, deslizándose hacia la nada con la misma facilidad con que se apaga una vela, para reavivarse más tarde, cuando las circunstancias son más propicias. ¿Es usted la misma persona cuyas aventuras de jardín de infancia recuerda (unas veces con gran claridad, otras con menos)? ¿Son las aventuras de este niño, cuya trayectoria en el espacio y en el tiempo corre aparentemente paralela a la trayectoria de su cuerpo, las mismas aventuras que las suyas? Este niño con su nombre, un niño cuya firma garabateada en un dibujo a lápiz le recuerda la manera en que *usted* solía firmar con su nombre, ¿es (era) usted este niño? El filósofo Derek Parfit (1984) ha comparado una persona con un club, una construcción humana muy diferente, que puede existir durante un año, y puede volver a constituirse años después con algunos de sus miembros (¿originales?). ¿Sería el mismo club? Lo sería si, por ejemplo, el club contemplara en sus estatutos la posibilidad de estos lapsus en su existencia. Pero podría no haber manera de decidir. Podría ser que conociéramos todos los hechos que pudieran tener alguna relación con el asunto y aun así ver que resultan inconcluyentes en cuanto a la *identidad* del (¿nue-

* Se pierde en la traducción de esta pregunta un juego de palabras con *selfless* (desinteresada/sin yo) y *herself* (sí misma/su yo). [N. del T.]

vo?) club. En la concepción de los yos —o las personas— que está surgiendo aquí, ésta es la analogía correcta; los yos no son perlas de alma que llevan una existencia independiente, sino artefactos de los procesos sociales que nos crean, y, como los demás artefactos, sujetos a repentinos cambios de estatuto. El único «impulso» que actúa en beneficio de la trayectoria de un yo, o de un club, es la estabilidad que le da el tejido de creencias que lo conforman, y cuando estas creencias pasan, pasa el yo, permanente o temporalmente.

Es importante no olvidar, puestos a considerar los ejemplos favoritos de los filósofos, el tan discutido fenómeno de los pacientes con el cerebro dividido. Un cerebro dividido es el resultado de una operación de *comisurotomía* en la que se ha cortado el *corpus callosum*, una ancha banda de fibras que conecta directamente los hemisferios izquierdo y derecho del córtex. La intervención deja los hemisferios conectados sólo de forma indirecta, a través de una serie de estructuras en el diencéfalo, por lo que es un procedimiento bastante drástico, que no debe adoptarse a menos que no queden alternativas. Alivia algunos casos de epilepsia para los que no queda otro tratamiento, al impedir las tormentas eléctricas generadas internamente que hacen que las crisis viajen por el córtex desde el «foco» de origen en uno de los hemisferios hasta el lado opuesto. La mitología filosófica sostiene que los pacientes con el cerebro dividido mostrarían una «división en dos yos» sin padecer por ello ninguna disminución seria de sus capacidades. La versión más atractiva de esta idea es que esas dos «caras» de la persona —el atento y analítico hemisferio izquierdo, y el tranquilo, intuitivo y holístico hemisferio derecho— se liberan durante la etapa postoperatoria, para brillar con más individualidad, ahora que el trabajo en equipo debe ser sustituido por una *détente* menos íntima. Es en verdad una idea atractiva, pero no es más que una exageración de la evidencia empírica que la inspiró. De hecho, sólo en una mínima fracción de casos se observa alguno de esos síntomas de identidad múltiple teóricamente tan sorprendentes. (Véanse, por ejemplo, Kinsbourne, 1974; Kinsbourne y Smith, 1974; Levy y Trevarthen, 1976; Gazzaniga y LeDoux, 1978; Gazzaniga, 1985; Oakley, 1985; Dennett, 1985b.)

No debe sorprendernos que los pacientes con el cerebro dividido, como los pacientes con visión ciega o con Trastornos de Personalidad Múltiple, no estén a la altura de lo que la filosofía espera de ellos, y eso no es culpa de nadie. Y no se debe al hecho de que los filósofos (y muchos otros intérpretes, incluidos los investigadores principales) exageren deliberadamente sus descripciones de estos fenómenos. Al contrario, en su esfuerzo por describir los fenómenos con la mayor concisión, se encuentran con que los recursos del lenguaje ordinario les empujan inexorablemente hacia el modelo del jefe del cuerpo, del espíritu en la máquina y de la audiencia en el Teatro Cartesiano. Nicholas Humphrey y yo, al comparar nuestras cuidadosas notas tomadas durante varias sesiones con pacientes que sufrían de TPM, descubrimos que habíamos caído, muy a pesar nuestro, en giros muy naturales

pero erróneos al describir lo que habíamos visto. Thomas Nagel (1971), el primer filósofo en escribir sobre los pacientes con el cerebro dividido, presentó una juiciosa y precisa relación de los fenómenos, tal como se los comprendía en aquel momento, y, reconociendo las dificultades de dar una explicación coherente, conjeturó: «Puede que nos resulte imposible abandonar ciertas maneras de concebirnos y de representarnos a nosotros mismos, por escasa que sea la evidencia en su favor proveniente de la investigación científica» (1971, pág. 397).

Ciertamente es difícil, pero no imposible. El pesimismo de Nagel es también exagerado. ¿Acaso no hemos conseguido liberarnos de esa manera tradicional de pensar? Quizás algunos no *quieran* abandonar la visión tradicional. Incluso puede haber buenas razones —razones morales— para intentar conservar el mito de los yos como perlas cerebrales, cosas particulares, concretas y contables, en vez de abstracciones, y para no querer tolerar la posibilidad de los cuasiyos, los semiyos y los yos de transición. Pero no cabe duda de que ésta es la manera correcta de interpretar el fenómeno de los cerebros divididos. Durante breves períodos de tiempo, en procedimientos experimentales cuidadosamente diseñados, algunos de estos pacientes se bifurcan como respuesta a una situación difícil, creando de forma temporal un segundo centro de gravedad narrativa. Puede que algunos efectos de esta bifurcación permanezcan indefinidamente bajo la forma de huellas en la memoria mutuamente inaccesibles, pero aparte de estas primitivas huellas de la bifurcación, la vida de un segundo y rudimentario yo sólo dura unos minutos como máximo, no lo suficiente como para derivar el tipo de autobiografía de que están compuestos los yos hechos y derechos. (Esto es también válido para la mayoría de las docenas de yos fragmentarios que desarrollan los pacientes con TPM; no hay bastantes horas en un día para que puedan despertarse y acumular más de algunos minutos de biografía exclusiva por semana.)

El carácter distintivo de las diversas narraciones es el elixir de la vida de los distintos yos. Como señala el filósofo Ronald de Sousa (1976):

> Cuando el doctor Jekyll se convierte en mister Hyde, estamos ante un asunto extraño y misterioso. ¿Son dos personas turnándose en ocupar un único cuerpo? Pero tenemos un caso aún más extraño: el doctor Juggle y el doctor Boggle, también se turnan en ocupar un único cuerpo, *¡pero son iguales como gemelos idénticos!* Pero, entonces, replicará usted, ¿por qué dicen que se han cambiado el uno por el otro? Bueno, ¿y por qué no?: si el doctor Jekyll puede convertirse en alguien tan diferente como Hyde, sin duda debe ser mucho *más fácil* para Juggle convertirse en Boggle, que es exactamente igual que él.
>
> Necesitamos un conflicto o una diferencia radical para hacer tambalear nuestro supuesto natural de que a un cuerpo le corresponde un solo agente como máximo (pág. 219).

¿*Qué se siente* al ser el hemisferio derecho de un paciente con el cerebro dividido? Ésta es la pregunta más natural del mundo,[1] y que conjura la imagen de sobresalto —y enfriamiento— mental: ahí está usted, atrapado en el hemisferio derecho de un cuerpo cuya parte izquierda usted conoce perfectamente (y todavía controla) y cuya parte derecha le resulta ahora tan remota como el cuerpo del primero que pase. A usted le gustaría contarle al mundo lo que se siente al ser usted, ¡pero no puede! Está usted aislado sin sus canales de comunicación verbal al haber perdido la línea directa con la emisora de radio en el hemisferio izquierdo. Usted hace lo que puede para hacer notar su presencia al mundo exterior, forzando su mitad de la cara en medias sonrisas y fruncimientos de ceño y, ocasionalmente (si es que usted es un yo de hemisferio derecho), garabateando una o dos palabras con su mano izquierda.

Este ejercicio de imaginación podría proseguir de forma evidente, pero sabemos que es una fantasía, tanto como las encantadoras historias de Beatrix Potter sobre Peter Rabbit y sus amigos animales antropomórficos. Y no lo es porque «la conciencia está sólo en el hemisferio izquierdo» ni porque *no pudiera ser* que alguien se encontrara en esa tesitura, sino simplemente porque *no es cierto* que la comisurotomía deje al despertar unas organizaciones lo bastante distintas y robustas como para mantener un yo separado.

Apenas sería un problema para mi teoría del yo el que fuese «lógicamente posible» la existencia en un paciente con el cerebro dividido de un hemisferio derecho como el descrito, pues mi teoría dice que no lo hay, y dice por qué: las condiciones para acumular el tipo de riqueza narrativa (y la independencia) que conforma un yo «hecho y derecho» no se dan. Mi teoría es igualmente impermeable a la afirmación —que nunca me atrevería a negar— de que *podría haber* conejitos que hablan, arañas que escriben mensajes en inglés sobre sus telas y, para el caso, trenecitos de vapor melancólicos. Supongo que podrían existir, pero no existen, así que mi teoría no tiene por qué explicarlos.

3. La insoportable levedad del ser

> Cualquier cosa que suceda, en cualquier momento o lugar, tendemos a preguntarnos qué o quién es el responsable. Esto nos lleva a descubrir explicaciones que de otro modo tal vez no imaginaríamos, y eso nos ayuda a predecir y controlar, no solamente lo que ocurre en el mundo, sino también lo que tiene lugar en nuestra mente. Pero, ¿qué ocurriría si esas mismas tendencias nos indujeran a imaginar cosas y causas inexistentes? Entonces inventaremos dioses falsos y supersticiones, y veremos su mano en cualquier coincidencia

1. Es interesante observar que Nagel, en 1971, ya se enfrentaba explícitamente a esta pregunta (pág. 398), antes de derivar su atención hacia los murciélagos; un asunto que trataremos en el capítulo siguiente.

casual. En realidad, quizás esa extraña palabra «yo» —como cuando decimos «yo tuve una buena idea»— refleja precisamente esa tendencia. Si nos sentimos obligados a encontrar alguna causa que explique todo lo que hacemos, pues bien, esa causa necesita un nombre. Usted la llama «yo». Yo la llamo «usted».

MARVIN MINSKY *The Society of Mind*, 1985, pág. 232.*

Un yo, de acuerdo con mi teoría, no es un viejo punto matemático, sino una abstracción que se define por la multitud de atribuciones e interpretaciones (incluidas las autoatribuciones y las autorrepresentaciones) que han compuesto la biografía del cuerpo viviente del cual es su centro de gravedad narrativa. Como tal, juega un papel particularmente importante en la economía cognitiva en curso de ese cuerpo viviente, porque, de todas las cosas del entorno sobre las cuales un cuerpo activo debe construir modelos mentales, ninguno es tan importante como el modelo que el agente tiene de sí mismo. (Véanse, por ejemplo, Johnson-Laird, 1988; Perlis, 1991.)

Para empezar, todo agente tiene que saber qué tipo de cosa es en el mundo. Esto, en un principio, puede parecer o trivial o imposible. «¡Yo soy yo!» no es una afirmación particularmente informativa, pero, ¿qué otra cosa puede necesitar saber uno, o qué otra cosa puede descubrir si todavía la desconoce? Para los organismos más simples, realmente es cierto que no hay mucho que decir sobre el autoconocimiento aparte del rudimentario saber biológico que encierran máximas tales como «¡Cuando tengas hambre, no te comas a ti mismo!» y «¡Cuando hay dolor, es tuyo!» En todo organismo, incluidos los seres humanos, el reconocimiento de estos principios básicos de diseño está «preconfigurado», forma parte del diseño subyacente del sistema nervioso, como el parpadeo cuando algo se acerca a los ojos o como el tiritar cuando hace frío. Una langosta podría comerse las pinzas de otra langosta, pero la perspectiva de comerse sus propias pinzas se ha hecho convenientemente impensable para ella. Sus opciones son limitadas, y cuando «piensa» en mover una pinza, su «pensador» está directa y apropiadamente conectado a la pinza que está pensando en mover. Con los seres humanos (y con los chimpancés y quizá con algunas otras especies), por otra parte, hay más opciones y, por tanto, más fuentes de confusión posibles.

Hace algunos años, las autoridades del puerto de Nueva York experimentaron un sistema de radar común para propietarios de pequeñas embarcaciones. Una potente antena de radar situada en tierra formaba una imagen radar del puerto, que podía transmitirse como una señal de televisión a los propietarios de los botes que, así, se podían ahorrar los costes de instalación de un radar, colocando un televisor en sus embarcaciones. ¿Cuál era la utilidad del sistema? Si usted estaba perdido en la niebla, y miraba la pantalla del televisor, veía que uno de esos puntos móviles en la pantalla era usted; pero, ¿cuál de ellos? He aquí un caso en que la pregunta, «¿qué cosa

* Página 240 de la edición castellana (Galápago, 1986). [N. del T.]

del mundo soy yo?», ni es trivial ni es imposible de responder. El misterio se desvela recurriendo a un simple truco: haga girar en círculo rápidamente su barco; entonces su punto es aquel que traza una pequeña «O» en la pantalla, a menos que muchos barcos decidan llevar a cabo la misma operación al mismo tiempo.

El método no es infalible, pero funciona la mayoría de las veces, e ilustra perfectamente un asunto mucho más general: a fin de controlar las complejas actividades en que se ocupa el cuerpo humano, el sistema de control del cuerpo (situado en el cerebro) debe ser capaz de reconocer una amplia variedad de información de entrada en tanto que datos sobre sí mismo, y cuando se produce algún dilema o hay motivos para el escepticismo, el único método fiable (aunque no infalible) de clasificar y asignar correctamente esa información consiste en efectuar pequeños experimentos: hacer algo y ver qué es lo que se mueve.[2] Un chimpancé puede aprender fácilmente a alcanzar unos plátanos a través de un agujero de su jaula, guiándose con la ayuda de un circuito cerrado de televisión, observando los movimientos de su brazo que aparecen en el monitor, instalado a una cierta distancia del brazo (Menzel y otros, 1985). Este acto de autorreconocimiento no es trivial, pues depende de ser capaz de percibir la consonancia existente entre los movimientos del brazo que se ven a través de la pantalla y los movimientos no vistos *pero intencionados* del brazo. ¿Qué ocurriría si los investigadores introdujeran un breve retraso en la cinta? ¿Cuánto tiempo cree usted que tardaría en darse cuenta de que lo que está viendo es su propio brazo (sin pistas verbales derivadas de la organización del experimento), si se introdujera un retraso de, por ejemplo, veinte segundos en el circuito cerrado?

La necesidad de autoconocimiento va más allá del problema de identificar los signos externos de nuestro movimiento corporal. Necesitamos saber acerca de nuestros estados internos, tendencias, decisiones, puntos fuertes y puntos débiles, y el método básico para obtener este conocimiento es esencialmente el mismo: hacer algo y «mirar» a ver qué se «mueve». Un agente avanzado puede desarrollar prácticas que le sirvan para mantener un seguimiento de sus circunstancias tanto corporales como «mentales». En los seres humanos, como hemos visto, dichas prácticas suelen comportar incesantes turnos de relación de historias y de verificación de historias, algunas verdaderas y otras ficticias. Los niños lo hacen en voz alta (piénsese en Snoopy, cuando se dice a sí mismo mientras está sentado sobre el techo de su casita: «Aquí está el famoso as de la Primera Guerra Mundial...»). Los adultos lo hacemos de forma más elegante: en silencio, tácitamente, siguiendo sin es-

2. ¿Y cómo sabemos que somos nosotros quienes estamos haciendo algo? ¿Dónde obtenemos el autoconocimiento inicial necesario para aprovechar esta ventaja? Aparentemente, ésta ha sido una pregunta de una importancia fundamental para algunos filósofos (Castañeda, 1967, 1968; Lewis, 1979; Perry, 1979), lo que ha generado una bibliografía de una complejidad extrema. Si éste es un problema filosófico sustancial, entonces debe de haber algo malo en la respuesta «trivial» (pero no sé qué): obtenemos nuestro autoconocimiento original básico de la misma manera que lo obtiene la langosta; estamos configurados de esta manera.

fuerzo la diferencia entre nuestras fantasías y nuestras reflexiones «serias». El filósofo Kendall Walton (1973, 1978) y el psicólogo Nicholas Humphrey (1986) han demostrado desde diferentes perspectivas la importancia de la representación, del contar historias y del fenómeno más fundamental de la simulación como elementos fundamentales para que practiquen los seres humanos que aún son unos tejedores de yos novatos.

Así pues, construimos una historia que nos define a nosotros mismos, organizada alrededor de una especie de punto luminoso de autorrepresentación (Dennett, 1981a). El punto no es un yo, por supuesto; es una *representación* de un yo (y el punto en la pantalla de radar para la isla de Ellis representación de una isla). Lo que hace que un punto sea el punto-*yo* y que otro punto sea el punto-*ella* el punto-*él* o el punto-*ello* no es la forma en que aparece, sino aquello para lo que se utiliza. Recoge y organiza la información sobre el yo del mismo modo que otras estructuras en mi cerebro hacen un seguimiento de la información sobre Boston, sobre Reagan o sobre los helados.

¿Y dónde está aquello *sobre* lo que trata su autorrepresentación? Está allí donde esté usted (Dennett, 1978b). ¿Y *qué* es esa cosa? Pues no es nada más, ni nada menos, que su centro de gravedad narrativa.

Otto vuelve:

> «El problema con los centros de gravedad es que no son reales; son las ficciones de un teórico.»

Éste no es el problema de los centros de gravedad; es su gloria. Son unas ficciones *magníficas*, unas ficciones que cualquiera se sentiría orgulloso de haber creado. Y los personajes ficticios de la literatura son aún mejores. Piense en Ishmael, de *Moby-Dick*. «Llámenme Ishmael», así es como empieza la novela, y nosotros obedecemos. No llamamos Ishmael al texto, ni llamamos Ishmael a Melville. ¿A quién o a qué llamamos Ishmael? Llamamos Ishmael a Ishmael, el maravilloso personaje ficticio que hallamos en las páginas de *Moby-Dick*. «Llámenme Dan» oyen salir de mis labios, y ustedes obedecen, pero no llamando Dan a mis labios, o a mi cuerpo, sino llamándo*me* Dan *a mí*, la ficción de un teórico creada por... bueno, no por mí, sino por mi cerebro, actuando en concierto durante años con mis padres, hermanos y amigos.

> «Eso estará muy bien para usted, pero yo soy completamente real. Puede que yo haya sido creado por los procesos sociales a los que usted acaba de aludir (así debe de haber sido, a no ser que ya existiera antes de mi nacimiento), ¡pero lo que esos procesos crearon es un yo *real*, no un mero personaje ficticio!»

Me parece que ya veo adónde quiere usted llegar. Si el yo no es algo real, entonces ¿qué ocurre con la responsabilidad moral? Una de las funciones

principales de un yo en nuestro esquema conceptual tradicional es la de ser aquello a lo que hay que cargarle el muerto, según rezaba el cartel de Harry Truman. Si los yos no son reales —no son *realmente* reales—, ¿acaso no nos iremos pasando el muerto los unos a los otros para siempre? Si no hay un Despacho Oval en el cerebro, donde trabaja una autoridad superior a quien se le puedan atribuir las decisiones, parece que estamos amenazados por una kafkiana burocracia de homúnculos, que, cuando se les pregunta, siempre contestan lo mismo: «Yo no sé nada, yo sólo trabajo aquí». La tarea de construir un yo capaz de *asumir* responsabilidades es un proyecto social y educativo muy importante, y tiene usted razón al preocuparse ante cualquier cosa que pueda amenazar su integridad. Pero una perla cerebral, un lo-que-sea real e «intrínsecamente responsable», no es más que una patética chucheria, como un amuleto de la suerte, con que combatir esta amenaza. La única esperanza es llegar a comprender, de forma naturalista, de qué manera desarrollan los cerebros sus autorrepresentaciones, equipando, así, cuando todo va bien, los cuerpos que controlan con unos yos responsables; y no es ésta una empresa desesperada. El libre albedrío y la responsabilidad moral son cosas que merece la pena perseguir, y, como he intentado demostrar en mi libro *Elbow Room: The Varieties of Free Will Worth Wanting* (1984), su mejor defensa consiste en abandonar el mito desesperado y plagado de contradicciones de la existencia de un alma distinta y separada.

«Pero, entonces, ¿yo no existo?»

Por supuesto que sí. Ahí está usted, sentado en una silla, leyendo mi libro y planteándome sus críticas. Y curiosamente su actual encarnamiento, pese a ser una condición previa necesaria para su creación, no es un requisito obligado para que su existencia pueda prolongarse de forma indefinida. Si usted fuese una alma, una perla de sustancia inmaterial, sólo podríamos «explicar» su potencial inmortalidad postulándola como una propiedad inexplicable, una *virtus dormitiva* que no se puede eliminar de la sustancia-alma. Y si usted fuese una perla de sustancia material, algún grupo de átomos de su cerebro espectacularmente especial, su mortalidad dependería de las fuerzas físicas que los mantienen unidos (podríamos preguntar a los físicos en qué consiste la «media-vida» de un yo). Si usted piensa en usted como un centro de gravedad narrativa, por otra parte, su existencia depende de la persistencia de esa narración (un poco como *Las mil y una noches*, pero en un único cuento), que *en teoría* podría sobrevivir a una serie indefinida de cambios de *medio*, podría teletransportarse fácilmente (en principio) como el noticiario de la noche y almacenarse por tiempo indefinido en forma de mera información. Si lo que usted es, es esa organización de la información que ha estructurado el sistema de control de su cuerpo (o, por plantearlo de manera más provocativa y, a la vez, más usual, si lo que usted es, es el programa que corre en el ordenador de su cerebro), entonces, en principio, usted podría sobrevivir a la muerte de su cuerpo tan intacto como

un programa que puede sobrevivir a la destrucción del ordenador en el que fue creado por primera vez. Algunos pensadores (por ejemplo, Penrose, 1989) consideran que ésta es una implicación espantosa y profundamente contraintuitiva de la visión que he defendido aquí. Pero si lo que usted está buscando es la inmortalidad, las alternativas son, simplemente, indefendibles.

Capítulo 14

LA CONCIENCIA IMAGINADA

1. Imaginando un robot consciente

En los capítulos anteriores hemos explicado los fenómenos de la conciencia humana en términos de operaciones de una «máquina virtual», una especie de programa de ordenador evolucionado (y en evolución) que coforma las actividades del cerebro. No hay Teatro Cartesiano; sólo hay Versiones Múltiples, compuestas por procesos de fijación de contenido que juegan diversos papeles semiindependientes dentro de la economía cerebral de controlar la singladura de un cuerpo humano por la vida. La sorprendente e insistente convicción de que hay un Teatro Cartesiano es el resultado de una serie de ilusiones cognitivas que hemos expuesto y explicado. Los «qualia» han quedado sustituidos por estados disposicionales complejos del cerebro, y el yo (también conocido con los nombres de audiencia en el Teatro Cartesiano, Significador Central o testigo) resulta ser una valiosa abstracción, la ficción de un teórico en vez de un observador o un jefe interno.

Si el yo no es «más» que el centro de gravedad narrativa, y si todos los fenómenos de la conciencia humana son explicables «solamente» como las actividades de una máquina virtual realizada en las conexiones astronómicamente ajustables de un cerebro humano, entonces, en principio, un robot «programado» de forma adecuada, con un cerebro basado en la química del silicio, podría ser consciente, podría tener un yo. Mejor dicho, podría existir un yo consciente cuyo cuerpo sería el robot y cuyo cerebro sería el ordenador. Esta implicación de mi teoría, para algunas personas es algo obvio y contra lo que no hay nada que objetar. «*¡Claro* que somos máquinas! Somos máquinas muy, muy complicadas y muy, muy evolucionadas, hechas de moléculas orgánicas en vez de metal y silicio, y nosotros somos conscientes, de modo que también puede haber máquinas conscientes: nosotros.» Para estos lectores, esta implicación era una conclusión anunciada. Lo que puede haber sido interesante para ellos, espero, es la variada serie de implicaciones nada evidentes que hemos encontrado por el camino, en particular aquellas que demuestran hasta qué punto la imagen cartesiana, a la que apela nuestro sentido común, debe ir siendo sustituida a medida que aprendemos más sobre los mecanismos cerebrales.

Otras personas, sin embargo, consideran la idea de que pueda haber, en

principio, robots conscientes como algo tan increíble que, para ellos, equivale a una *reductio ad absurdum* de mi teoría. En una ocasión, un amigo mío respondió a mi teoría con una sincera confesión: «Pero, Dan, ¡no puedo imaginar un robot consciente!». Es probable que algunos lectores se sientan inclinados a aceptar esta afirmación, pero deberían intentar resistir la tentación de aceptarla, porque mi amigo cometió un error. Su error fue muy simple, pero llama la atención sobre una confusión fundamental que impide que progresemos en nuestra comprensión de la conciencia. «*Sabes* perfectamente que eso que dices es falso», le contesté, «más de una vez has imaginado un robot consciente. El problema no es que no puedas imaginar un robot consciente, sino que no puedes imaginar *cómo* puede ser consciente un robot.»

Cualquiera que haya visto a R2D2 o a C3PO en *La guerra de las galaxias*, o escuchado a Hal en *2001*, ha imaginado un robot consciente (o un ordenador consciente; que el sistema tenga una total autonomía, como R2D2, o permanezca fijo en algún lugar, como Hal, no es realmente importante para nuestro ejercicio de imaginación). La verdad es que imaginar el flujo de la conciencia de una cosa «inanimada» es como un juego de niños. Los niños lo hacen constantemente. No sólo los ositos de peluche poseen una vida interior; también la tiene la pequeña locomotora. En el bosque, los abetos esperan en silencio, temerosos del hacha del leñador y a la vez deseando convertirse en un hermoso árbol de navidad, en una casa calentita y rodeados de niños. La literatura infantil (por no hablar de la televisión) está repleta de oportunidades para imaginar la vida consciente de las cosas. Los artistas que ilustran estas fantasías suelen ayudar a la imaginación de los niños pintando unas caras muy expresivas a todos estos agentes imaginarios, pero esto no es esencial. Con que hablen —como lo hace Hal— basta, sin la necesidad de que haya una cara, para fomentar la ilusión de que hay alguien ahí dentro, de que se puede sentir algo al ser Hal, o un osito de peluche, o una vieja locomotora.

Ahí está el problema, claro: esto no son más que ilusiones, o, por lo menos, eso es lo que parece. Hay diferencias entre la una y la otra. Es evidente que los ositos de peluche no son conscientes, pero no es tan evidente que un robot no pueda serlo. Lo que es evidente es que resulta difícil imaginar cómo podría serlo. Mi amigo, al no poder imaginar *cómo* podría ser consciente un robot, se resistió a imaginar un robot que *fuese* consciente, aunque lo podría haber hecho sin la mayor dificultad. Hay una enorme diferencia entre estos dos ejercicios de imaginación, pero la gente tiende a confundirlos. Es terriblemente difícil, sin lugar a dudas, imaginar de qué modo el ordenador-cerebro de un robot puede ser el soporte de una conciencia. ¿Cómo es *posible* que una compleja serie de acontecimientos de procesamiento de la información ejecutados en el interior de un puñado de *chips* de silicio equivalga a una serie de experiencias conscientes? Pero eso es tan difícil como imaginar de qué modo un cerebro humano orgánico puede ser el soporte de una conciencia. ¿Cómo es *posible* que una compleja serie de interaccio-

nes electroquímicas entre miles de millones de neuronas equivalga a una serie de experiencias conscientes? Y sin embargo, no tenemos ninguna dificultad en imaginar a seres humanos conscientes, aunque todavía no podamos imaginar *cómo* puede ser esto posible.

¿Cómo puede ser el cerebro la base de la conciencia? Por lo general, los filósofos han tratado esta pregunta como una cuestión retórica, sugiriendo implícitamente que la respuesta está muy por encima de la capacidad de comprensión de los humanos. Uno de los principales objetivos de este libro ha sido el de derribar este supuesto. He argumentado que es *posible* imaginar de qué modo esa serie de actividades en el cerebro equivalen a experiencias conscientes. Mi argumento es muy simple: me he limitado a mostrar cómo hacerlo. Resulta que la manera de imaginarlo consiste en pensar en el cerebro como si fuera una especie de ordenador. Los conceptos de la informática y las ciencias de la computación nos proporcionan los elementos necesarios para imaginar, si queremos cruzar esa *terra incognita* que se extiende entre nuestra fenomenología, tal como la conocemos por «introspección», y nuestros cerebros, tal como nos los presenta la ciencia actual. Al pensar en nuestros cerebros como sistemas de procesamiento de la información, poco a poco podemos ir disipando la niebla y hallar así el camino para cruzar el gran abismo, desvelando de qué modo nuestros cerebros producen todos esos fenómenos. Hay muchos obstáculos y muchas trampas que debemos evitar —tentadores callejones sin salida tales como el Significador Central, la «repleción» y los «qualia», por ejemplo—, y no cabe duda de que todavía permanecen algunas confusiones y algunos errores de envergadura en el esbozo que he presentado, pero, cuando menos, ahora podemos ver cómo será nuestro camino.

No obstante, algunos filósofos han afirmado que es totalmente imposible cruzar ese abismo. Thomas Nagel (1974, 1986) ha sostenido que no hay manera de pasar al nivel subjetivo de la fenomenología partiendo del nivel objetivo de la fisiología. Y más recientemente, Colin McGinn ha afirmado que la conciencia posee una «estructura oculta» que se extiende más allá de la fenomenología y de la fisiología y que, aunque dicha estructura oculta nos permitiría tender un puente sobre el abismo, probablemente éste quedaría fuera de nuestro alcance para siempre.

> El tipo de estructura oculta que propongo no se situaría a ninguno de los niveles descritos por Nagel: se situaría en algún lugar entre ellos. Ni fenomenológico ni físico, este nivel intermedio no estaría cortado (por definición) según el patrón de los otros dos lados del abismo, y, por tanto, no le sería imposible alcanzar la otra orilla. Su caracterización requeriría una radical innovación conceptual (que, según he argumentado, superaría los límites de nuestras capacidades) (McGinn, 1991, págs. 102-103).

El nivel de descripción del *software* o de la «máquina virtual» que he descrito en este libro es precisamente el tipo de nivel que describe McGinn: por

una parte ni explícitamente fisiológico o mecánico y, aun así, capaz de tender los puentes necesarios hacia los mecanismos cerebrales y, por la otra parte, no explícitamente fenomenológico pero capaz de establecer los vínculos necesarios con el mundo del contenido, con los mundos de la (hetero-) fenomenología. ¡Lo hemos conseguido! *Hemos* imaginado cómo puede producir experiencias conscientes un cerebro. ¿Por qué McGinn piensa que está por encima de nuestras capacidades el embarcarnos en esta «radical innovación conceptual»? ¿Acaso ha sometido los diversos enfoques de la mente basados en la idea del *software* al detallado análisis que podría demostrar su futilidad? No. Ni siquiera los toma en consideración. Ni siquiera intenta imaginar el nivel intermedio cuya existencia él mismo ha postulado; se limita a señalar que a él le parece evidente que no hay ninguna esperanza.

Esta espuria «obviedad» es un gran obstáculo para el progreso de la comprensión del fenómeno de la conciencia. Lo más natural del mundo es pensar que la conciencia es algo que se produce en una especie de Teatro Cartesiano, y suponer que nada tiene de malo el pensar de esta manera. Todo eso parece obvio hasta que uno intenta aprender en serio las actividades del cerebro, y empieza a intentar imaginar, con detalle, un modelo alternativo. Entonces lo que ocurre se parece mucho a descubrir el truco de un juego de manos. Cuando nos paramos a mirar entre bastidores, descubrimos que no vimos realmente lo que creíamos haber visto en el escenario. El enorme abismo que se extiende entre la fenomenología y la fisiología empieza a cerrarse; comprobamos que algunas de las características más «obvias» de la fenomenología no son reales: no hay repleción con figmento; no hay qualia intrínsecos; no existe el manantial del significado y la acción; no existe ese mágico lugar donde se produce el entendimiento. En efecto, no hay ningún Teatro Cartesiano; incluso la distinción entre experiencias sobre el escenario y procesos entre bastidores empieza a perder todo su atractivo. Todavía nos quedan muchos fenómenos sorprendentes por explicar, pero algunos de los efectos especiales que más aturdían a nuestra mente ni siquiera existen, así que no tenemos que tomarnos la molestia de explicarlos.

Una vez hemos hecho algunos procesos en esta difícil tarea de imaginar cómo produce el cerebro los fenómenos de la conciencia, debemos hacer algún ajuste en la tarea más fácil: imaginar a alguien o algo que es consciente. Podemos seguir pensando en ello apelando a una especie de flujo de la conciencia, pero ya no podemos dotar a ese flujo con todas sus propiedades tradicionales. Ahora el flujo de la conciencia ha sido reconcebido como las operaciones de una máquina virtual realizada en el cerebro, ya no resulta «obvio» el que hayamos sucumbido a una ilusión cuando imaginamos ese flujo produciéndose en el ordenador cerebral de un robot, por ejemplo.

McGinn invita a sus lectores a unírsele en la rendición: es totalmente imposible imaginar cómo puede el *software* hacer consciente a un robot. Ni siquiera lo intenten, dice. Otros filósofos han fomentado esta actitud recurriendo a experimentos mentales que «funcionan» precisamente porque disuaden al lector de llevar a cabo el intento de imaginar, con detalle, cómo

puede el *software* llegar a conseguirlo. Curiosamente, los dos experimentos más conocidos tienen que ver con China: el experimento de la Nación China de Ned Block (1978) y el de la Habitación China de John Searle (1980, 1982, 1984, 1988a, b).[1] Ambos experimentos mentales se basan en el mismo y mal dirigido ejercicio de imaginación, y puesto que el de Searle ha sido el más ampliamente discutido, me centraré exclusivamente en él. Searle nos invita a imaginarle encerrado en una habitación, simulando de forma manual un programa de IA gigante que, supuestamente, comprende el chino. Searle estipula que el programa es capaz de superar el test de Turing, frustrando así todo intento por parte de sus interlocutores humanos de distinguirlo de un sistema realmente capaz de comprender el chino. De esta mera imposibilidad de distinguirlo desde el punto de vista de la conducta no se deduce, afirma Searle, que, en la Habitación China, se produzca una comprensión real del chino o una conciencia china. Searle, encerrado en la habitación y manipulando concienzudamente las cadenas de símbolos del programa de acuerdo con éste, nunca llega a tener una comprensión del chino, y no hay nada más en la habitación que comprenda el chino (eso «es obvio» como diría Frank Jackson).

Se supone que este experimento mental prueba la imposibilidad de lo que Searle denomina «IA fuerte», la tesis de que «un ordenador digital programado adecuadamente y con la información de entrada y de salida apropiada tendría una mente exactamente en el mismo sentido que los seres humanos tienen una mente» (Searle, 1988a). Las diferentes versiones de este experimento mental que Searle ha ido produciendo en la última década han provocado gran cantidad de reacciones y, aunque filósofos y otras personas siempre han hallado defectos en su estructura como argumento lógico,[2] parece innegable que su «conclusión» sigue siendo «obvia» para muchas personas. ¿Por qué? Porque nadie se imagina realmente la situación con el detalle que ésta requiere.

He aquí un simple experimento que nos ayudará a demostrar que mi diagnóstico es el correcto. Primero, imaginemos un breve extracto del diálogo entre la Habitación China y el juez que dio un veredicto positivo en el test

1. Intenten imaginar cuál sería la impresión de Ji Hu-Min, uno de mis estudiantes de doctorado procedente de Beijing, cuya introducción a la filosofía de la mente angloamericana (cuando su inglés era todavía bastante rudimentario) consistió en asistir a un seminario donde los estudiantes y el profesor debatían con vehemencia sobre lo que ocurriría si toda la población China se viera obligada a participar en la ejecución masiva de un programa de IA presuntamente consciente (el ejemplo de Block), pasando inmediatamente a discutir, con la misma falta de sensibilidad hacia el observador procedente de China, el experimento de la Habitación China de Searle.

2. La refutación definitiva, y a la que Searle nunca ha respondido de forma adecuada, se debe a Douglas Hofstadter, en Hofstadter y Dennett (1981), págs. 373-382. A lo largo de los años han ido apareciendo muchas otras críticas. En «Fast thinking» (en Dennett, 1987a), yo mismo me ocupaba de hacer un nuevo diagnóstico del origen de los puntos más confusos de este experimento mental. La respuesta de Searle consistió en afirmar, sin ninguna argumentación que apoyara sus conclusiones, que todas mis observaciones son irrelevantes (Searle, 1988b). A ningún mago le gusta que se revelen sus trucos en público.

de Turing. (Para facilitar las cosas, me he permitido la libertad de traducirlo del chino al inglés.)*

> JUEZ: ¿Sabe usted aquel del irlandés que se encontró una lámpara maravillosa? Al frotarla, se le apareció un genio que le concedió tres deseos. «¡Me tomaría una pinta de Guinness!», replicó el irlandés, y ésta se le apareció al instante. Nuestro hombre se lanzó ávidamente sobre ella, bebiendo a grandes sorbos, pero el nivel de Guinness en la jarra siempre se reponía como por arte de magia. Al cabo de un rato, el genio empezó a impacientarse. «Bueno, ¿y qué pasa con el segundo deseo?», le preguntó. A lo que, entre sorbo y sorbo, el irlandés replicó, «Ah, sí, ¡pues me tomaría otra igual!».
>
> HABITACIÓN CHINA: Muy gracioso. No, no lo sabía, pero encuentro que los chistes racistas son de muy mal gusto, ¿sabe? Me he reído muy a pesar mío, pero, la verdad, esperaba que usted encontraría otros temas de discusión.
>
> J: Está bien, pero le conté el chiste porque quisiera que me lo explicara.
>
> HC: ¡Qué aburrido! Los chistes nunca deberían explicarse.
>
> J: Como quiera, pero ésta es mi pregunta para el test. ¿Podría usted explicarme por qué «funciona» el chiste?
>
> HC: Si insiste. Veamos, depende del supuesto de que la jarra que se vuelve a llenar por arte de magia seguirá llenándose para siempre, de modo que el irlandés tiene toda la cerveza que quiera. Apenas puede haber algún motivo por el cual querría tener otra igual, pero es tan tonto (ésta es la parte que a mí me molesta) o está tan borracho, que no se da cuenta y, sintiéndose tan contento porque su primer deseo se hizo realidad, pide una segunda ronda. Evidentemente, estos supuestos son falsos, sólo forman parte de la situación especial que rodea a todo acto de contar un chiste, en que suspendemos nuestra incredulidad en la magia, etc. De hecho, podríamos imaginar una continuación del chiste un poco más elaborada, según la cual nuestro irlandés «hiciera bien» al pedir un segundo deseo igual que el primero: a lo mejor está planeando organizar una gran fiesta, y una sola jarra no se volvería a llenar con la celeridad suficiente como para satisfacer a sus sedientos invitados (y ya sabemos que no vale la pena guardar la cerveza, ya que la cerveza pasada pierde sabor). Todos tendemos a no considerar todos estos aspectos cuando nos cuentan un chiste, lo cual explica en parte que los chistes funcionen. ¿Es suficiente con esto?

Esta conversación no es particularmente deslumbrante, pero supongamos que con ella bastó para engañar al juez. Ahora se nos invita a imaginar que es el programa que Searle está simulando manualmente quien compone todos estos discursos de la HC. ¿Difícil? Por supuesto, pero dado que Searle estipula que el programa supera el test de Turing, y dado que este nivel de virtuosismo conversacional sin duda estaría dentro de sus capacidades, a menos que intentemos imaginar toda la complejidad de un programa capaz de generar una conversación como aquélla, no estaremos siguiendo las instrucciones. Asimismo, es evidente que debemos imaginar que Searle no tiene ni el más mínimo indicio de lo que está haciendo en la Habitación China;

* Y yo lo he traducido del inglés al castellano. [N. del T.]

él solo ve ceros y unos que manipula de acuerdo con el programa. Es importante, por cierto, que Searle nos invite a imaginar que lo que él manipula son incomprensibles caracteres chinos en vez de ceros y unos, ya que ello puede llevarnos a suponer (sin fundamento) que el programa gigante funciona por un simple proceso de «hacer coincidir» los caracteres chinos de entrada con los caracteres chinos de salida. Ningún programa así funcionaría, evidentemente; ¿acaso los discursos en inglés de la HC «coinciden» con las preguntas del juez?

Un programa que realmente pudiera generar los discursos de la HC en respuesta a las preguntas del J tendría más o menos esta apariencia (visto desde el nivel de la máquina virtual, pero no desde el nivel más básico en que se encuentra Searle). Al analizar las primeras palabras, «¿Sabe usted aquél...?», se activarían algunos de los demonios para la detección de chistes del programa, lo cual pondría en funcionamiento toda una serie de estrategias para tratar con ficciones, lenguaje «con segundas», y cosas parecidas, de modo que al analizar las palabras «lámpara maravillosa», el programa ya habría dado una prioridad muy baja a todas las respuestas con la función de hacer notar que las lámparas maravillosas no existen. Se activaría una serie de marcos (Minsky, 1975) o guiones (Schank y Abelson, 1977) acerca de chistes con un genio, creando así diversas expectativas sobre posibles continuaciones, las cuales no se verían satisfechas, finalmente, a causa del desenlace de la historia, que invocaría un guión más mundano (el guión para «pedir una segunda ronda»), aunque lo inesperado de esta situación no se perdería en el programa. (...) Al mismo tiempo, los demonios sensibles a las connotaciones negativas de los chistes racistas también se verían alertados, lo cual motivaría las palabras pronunciadas por la HC en su primera respuesta. (...) Así, sucesivamente, con mucha más complejidad de la que yo he podido reflejar aquí en este pequeño esbozo.

Lo cierto es que cualquier programa capaz de mantener una conversación como la que reprodujimos más arriba dispondría de un sistema extraordinariamente flexible y con muchos niveles, rebosante de «conocimiento del mundo», de metaconocimiento y de metametaconocimiento sobre sus propias respuestas, las posibles respuestas de su interlocutor, sus propias «motivaciones», las motivaciones de su interlocutor y mucho, mucho más. Searle no niega que los programas puedan tener una estructura como ésta, claro está. Simplemente nos invita a no tenerlo en cuenta. Sin embargo, si tenemos que hacer un buen ejercicio de imaginación con este caso, no sólo tenemos que intentarlo: estamos obligados a imaginarnos que el programa que Searle está simulando manualmente tiene toda esa estructura y mucha más, si es que podemos. Pero, entonces, ya no es tan *obvio* que no se produzca la comprensión del chiste que se ha contado. *Quizá* los miles de millones de acciones ejecutadas por todas esas partes perfectamente estructuradas produzcan, después de todo, un genuino entendimiento en el sistema. Si su respuesta ante esta hipótesis es que no tiene ni la más remota idea de si habría entendimiento o no en un sistema tan complejo, ello es suficiente para

demostrar que el experimento mental de Searle depende, ilícitamente, de que usted imagine un caso demasiado simple, un caso irrelevante, y de que saque usted una conclusión «obvia».

Así es como se produce el engaño a la imaginación. Vemos claramente que si hubiera entendimiento en un sistema gigante como éste, no sería el entendimiento de Searle (pues él no es más que un engranaje dentro de la maquinaria, completamente ignorante del contexto de lo que está haciendo). También vemos claramente que no hay ni el más remoto indicio de entendimiento en cualquier pedazo de programa que sea lo bastante pequeño como para imaginarlo; sea lo que sea, no es más que un estúpido mecanismo que transforma cadenas de símbolos en otras cadenas de símbolos de acuerdo con una receta sintáctica o mecánica. Y es aquí donde hay que rescatar la premisa que se suprime de forma implícita: seguramente, *más de lo mismo,* no importa cuánto más, nunca conseguirá que se produzca un genuino entendimiento. Pero, ¿por qué debemos aceptar esto como cierto? Los dualistas cartesianos lo aceptarían, porque piensan que ni siquiera los cerebros humanos son capaces de generar entendimiento por sí mismos; según la visión cartesiana, se necesita de un alma inmaterial para producir el milagro del entendimiento. Si por otra parte nos consideramos materialistas, convencidos de que de un modo u otro nuestros cerebros son los únicos responsables de sí mismos, sin necesidad de ninguna ayuda milagrosa, debemos admitir que el verdadero entendimiento se produce por un proceso compuesto de interacciones entre una serie de subsistemas que por sí solos no poseen entendimiento. El argumento que empieza diciendo «este poquito de actividad cerebral no entiende el chino, y tampoco lo entiende este poco más...», está condenado a llegar a la conclusión no deseada de que ni siquiera la actividad de todo el cerebro es suficiente para comprender el chino. Es *muy difícil imaginar* cómo es posible que «más de lo mismo» pueda resultar en entendimiento, pero tenemos buenas razones para suponer que efectivamente es así, de modo que, en este caso, tenemos que redoblar nuestros esfuerzos, no abandonar.

¿Cómo podemos redoblar nuestros esfuerzos? Con la ayuda de algunos conceptos útiles: el concepto del *software* de nivel intermedio que fue diseñado por los informáticos precisamente para ayudarnos a seguir lo que en caso contrario serían las complejidades inimaginables de sistemas muy grandes. En los niveles intermedios vemos muchas entidades que son invisibles en niveles más microscópicos tales como los «demonios» a que aludíamos más arriba, a los cuales se les atribuye un módico cuasientendimiento. Así, no es tan difícil llegar a imaginarse cómo es posible que «más de lo mismo» pueda producir un genuino entendimiento. Todos estos demonios y las demás entidades están organizados en un sistema enorme, cuyas actividades se organizan alrededor de su propio centro de gravedad narrativa. Searle, trabajando en la Habitación China, no entiende el chino, pero no está solo en la habitación. También está el sistema, la HC, y es a *este* yo al que debemos atribuir la comprensión del chiste.

Esta respuesta al ejemplo de Searle es lo que él ha bautizado con el nombre de respuesta de los sistemas. Es la misma respuesta que los investigadores en IA han venido dando desde que se propuso por primera vez este experimento, hace ya más de una década, pero rara vez se ha sabido apreciar su valor fuera de los límites de la IA. ¿Por qué? Probablemente porque aquellos que no lo han sabido apreciar tampoco han aprendido a imaginar un sistema así. No pueden imaginar cómo es posible que el entendimiento sea una propiedad que emerge a partir una cantidad inmensa de cuasientendimientos distribuidos en un sistema muy grande. Sin duda no podrán si no lo intentan, pero ¿qué ayuda podemos prestarles en este ejercicio tan difícil? ¿Es acaso «hacer trampas» el pensar en el *software* como algo que se compone de homúnculos que casi comprenden, o es precisamente el tipo de ayuda que se necesita para que la imaginación pueda tratar con esa complejidad astronómica? Searle comete una falacia. Nos invita a imaginar que el programa gigante posee una arquitectura muy simple consistente en consultar una tabla y en emparejar secuencias de caracteres chinos con otras secuencias del mismo tipo, como si algo así pudiera ser considerado como un programa. No tenemos que molestarnos en imaginar un programa tan simple ni en asumir que *éste* es el programa que Searle está simulando, ya que ningún programa como éste sería capaz de producir los resultados que le permitirían superar el test de Turing, como se nos dice. (Para un argumento similar y su refutación, véase Block, 1982; y Dennett, 1985a.)

La complejidad es relevante. Si no lo fuera, dispondríamos de un argumento mucho más simple en contra de la IA fuerte: «Esta calculadora de bolsillo no entiende el chino, y cualquier ordenador que podamos concebir no es más que una calculadora gigante, de modo que ningún ordenador puede entender el chino. Q.E.D.». Cuando tenemos en cuenta la complejidad, como debe ser, tenemos que tenerla realmente en cuenta, y no sólo hacer ver que la estamos teniendo en cuenta. Esto es difícil de hacer, pero hasta que no lo hagamos, cualquier intuición que tengamos sobre lo que es «obvio» que no está presente no debe ser tenida en cuenta. Como en el caso del experimento que nos propone Frank Jackson sobre Mary, la investigadora del color, el experimento mental de Searle nos lleva a una fuerte y clara convicción sólo si no seguimos las instrucciones. Estas bombas de intuición son defectuosas; no sirven para dirigir nuestra imaginación, sino para desorientarla.

¿Qué ocurre entonces con mis propias bombas de intuición? ¿Qué decir de Shakey el robot, o del CADBLIND Mark II, o del paciente con visión ciega entrenado en el *biofeedback*, por ejemplo? ¿No son también sospechosos y culpables de desorientar al lector? He puesto mucho cuidado en contar estas historias a fin de dirigir su imaginación en ciertas direcciones y para evitar que usted se enzarzara con complejidades que me parecían innecesarias para presentar las ideas que pretendía ilustrar. Existe, no obstante, una asimetría: mis bombas de intuición pretenden, en gran medida, ayudarle a imaginar nuevas posibilidades, y no convencerle de que ciertas perspecti-

vas son imposibles. Hay excepciones. Mi variante sobre el tema del cerebro en un tarro con el que empezaba el libro fue diseñada con la idea de hacerle ver que ciertas formas de engaño son imposibles, y algunos de los experimentos mentales del capítulo 5 pretendían demostrar que, a menos que exista un Teatro Cartesiano, *no puede* existir un hecho determinante que permita distinguir una revisión de contenido de tipo orwelliano de otra de tipo estaliniano. Estos experimentos mentales procedían, sin embargo, clarificando primero la postura de la «oposición»; los ejemplos de la mujer con sombrero en la fiesta y la mujer del pelo largo con gafas, por ejemplo, se diseñaron para clarificar la intuición que después intenté desacreditar con mi argumentación.

Sin embargo, que sea precavido el lector: mis bombas de intuición, como las de cualquier otro, no son las simples demostraciones que parecen ser; tienen más de arte que de ciencia. (Para algunas advertencias más sobre los experimentos mentales de los filósofos, véase Wilkes, 1988.) Si nos ayudan a concebir nuevas posibilidades, que podamos confirmar después recurriendo a métodos más sistemáticos, entonces es un logro; si, por el contrario, hacen que nos adentremos por un incierto camino de rosas, entonces es una pena. Incluso las mejores herramientas pueden utilizarse mal, y como cualquier trabajador, mejor será que comprendamos bien cómo funcionan las nuestras.

2. Lo que se siente al ser un murciélago

El experimento mental sobre la conciencia más ampliamente citado y que más influencia ha ejercido en los más variados círculos es el experimento de Thomas Nagel sobre «¿Qué se siente al ser un murciélago?» («What is it like to be a bat?», 1974). Nagel responde a la pregunta del título afirmando que para nosotros es totalmente imposible imaginar algo así. Aparentemente, muchos simpatizan con esta conclusión; no es raro ver citado este artículo en trabajos científicos como si fuera una gran rareza, un «resultado» filosófico, la demostración de un hecho que toda teoría debe incorporar.

Nagel supo escoger muy bien las criaturas que iban a ser objeto de su reflexión. Los murciélagos, como mamíferos, se parecen lo bastante a nosotros como para que exista la convicción de que *sin duda* son conscientes. (Si hubiera escrito «¿Qué se siente al ser una araña?», muchos se habrían sentido con el derecho de preguntarle qué le hacía pensar que realmente se sentía algo al ser una araña.) Pero gracias a su sistema de ecolocalización —los murciélagos pueden «ver con sus oídos»— también son lo bastante diferentes de nosotros como para que todos podamos percibir la enorme distancia que nos separa de ellos. Si hubiera escrito un artículo titulado «¿Qué se siente al ser un chimpancé?» o, mejor, «¿Qué se siente al ser un gato?», la opinión de que su resultado tan pesimista es obvio no habría sido tan unánime. Hay muchas personas que están completamente convencidas de que

saben exactamente lo que se siente al ser un gato. (Esas personas están equivocadas, por supuesto, a menos que hayan complementado todo su amor y toda su simpatía al observar a los gatos con buenas dosis de investigación psicológica, aunque, en cualquier caso, según Nagel, estarían pecando por exceso de confianza en estas herramientas de investigación.)

Para bien o para mal, la mayoría de las personas no tienen ningún problema en aceptar el «resultado» de Nagel en cuanto a nuestra incapacidad de acceder a la conciencia de un murciélago. Algunos filósofos, sin embargo, lo han puesto en duda, y con razón (Hofstadter, 1981b; Hardin, 1988; Leiber, 1988; Akins, 1990). En primer lugar, es preciso aclarar de qué resultado se trata. No es la afirmación epistemológica o empírica de que si alguien consiguiera («por accidente») imaginar lo que se siente al ser un murciélago, nunca sería capaz de confirmar que había conseguido realizar esa proeza imaginativa. Consiste, por el contrario, en decir que los seres humanos no disponemos, ni nunca podremos disponer, de los recursos, los mecanismos de representación, para representarnos a nosotros mismos lo que se siente al ser un murciélago.

La distinción es importante. En el capítulo 12 nos ejercitamos en realizar la proeza de imaginar lo que se debió sentir al ser un ciudadano de Leipzig escuchando una de las cantatas de Bach por primera vez. El problema epistemológico es complejo, pero fácil de abordar mediante los métodos normales de investigación. El imaginar exactamente qué tipo de experiencias habría tenido, y en qué medida diferirían de nuestras propias experiencias sobre Bach, es una mera cuestión de investigación histórica, cultural, psicológica y, quizá, fisiológica. Podemos imaginarnos algo de esto con cierta facilidad, incluidas algunas de las diferencias más sorprendentes con nuestra propia experiencia, pero si intentáramos colocarnos en la propia secuencia de estados de experiencia de los que disfrutaría una persona así, nuestras ganancias disminuirían notablemente. Una tarea de este calibre exigiría que sufriésemos transformaciones radicales, olvidando la mayor parte de nuestros conocimientos, perdiendo asociaciones y hábitos, y adquiriendo nuevos hábitos y asociaciones. Podemos utilizar nuestro método de investigación de la «tercera persona» para determinar cuáles deberían ser dichas transformaciones, pero someternos a ellas comportaría unos costes excesivos en cuanto a aislamiento de nuestra cultura contemporánea: sin escuchar la radio, ni leer sobre los desarrollos políticos y sociales posteriores a Bach, etc. No es necesario llegar a estos extremos para aprender sobre la conciencia de un ciudadano de Leipzig.

Lo mismo vale para imaginarse lo que se siente al ser un murciélago. Deberíamos interesarnos por lo que podemos aprender sobre la conciencia del murciélago (si es que la tiene), y no sobre si podemos convertir nuestras mentes temporal o permanentemente en mentes de murciélago. En el capítulo 12 socavamos el supuesto de que hay propiedades «intrínsecas» —los qualia— que constituyen *lo que se siente* al tener una u otra experiencia consciente y, como ha señalado Akins (1990), incluso si hubiera propiedades resi-

duales de experiencia de murciélago, no disposicionales y no relacionales, llegar a tener un conocimiento íntimo de éstas, aun permaneciendo ignorantes de los hechos investigables sobre la estructura sistemática de la conducta y la percepción del murciélago, nos dejaría en el más profundo desconocimiento sobre lo que se siente al ser un murciélago. Hay bastantes cosas que podemos llegar a conocer sobre lo que se siente al ser un murciélago, y ni Nagel ni ningún otro nos han demostrado que haya algo particularmente interesante o teóricamente relevante que nos sea inaccesible.

Nagel afirma que ningún conocimiento adquirido desde la perspectiva de la tercera persona podrá decirnos nada sobre lo que se siente al ser un murciélago, y yo niego rotundamente esta afirmación. ¿Cómo podemos dirimir esta disputa? Pues participando en algo que empieza como un juego de niños, un juego en el que una persona imagina lo que se siente al ser *x*, y la otra intenta demostrar que hay algo incorrecto en ese particular ejercicio de heterofenomenología.

Aquí tenemos unos simples ejercicios para ir entrando en calor:

> «*A*: Éste es Pooh, el osito de peluche, pensando que le gustaría desayunar un poco de miel.»
> «*B*: Mal. Un osito de peluche no tiene la capacidad de distinguir la miel de ninguna otra cosa. No tiene órganos sensoriales que funcionen, ni tampoco tiene estómago. El osito está relleno de materia inerte. No se siente nada al ser un osito de peluche.»
> «*A*: Éste es Bambi, el cervatillo, admirando una hermosa puesta de sol, hasta que, de repente, el tono anaranjado del cielo le recuerda la chaqueta del malvado cazador.»
> «*B*: Mal. Los ciervos son ciegos al color (bueno, quizá tengan una especie de visión dicromática). De entre todas las cosas de que los ciervos son conscientes (si es que son conscientes de algo), no se encuentra la posibilidad de distinguir colores como el naranja.»
> «*A*: Éste es Billy, el murciélago, percibiendo, con su especial sistema de sonar, que el objeto volador que está descendiendo hacia él no es su primo Bob, sino un águila con las alas extendidas y las garras prestas para atacar.»
> «*B*: Un momento, ¿a qué distancia has dicho que estaba el águila? La ecolocalización de los murciélagos sólo funciona a unos pocos metros.»
> «*A*: Hmmm, bueno, pues... ¡el águila está a tan sólo dos metros!»
> «*B*: Ah, pues ahora resulta difícil decirlo. ¿Cuáles son los límites de resolución exactos de la ecolocalización de los murciélagos? ¿Se utiliza para identificar objetos, o sólo como sistema de alerta y de seguimiento para cazar? ¿Sería capaz un murciélago de distinguir unas alas extendidas de unas alas cerradas mediante la ecolocalización? Lo dudo, pero tendremos que diseñar nuevos experimentos para descubrirlo y también, claro está, algunos experimentos para descubrir si los murciélagos son capaces de seguir, e identificar, a sus parientes. Algunos

mamíferos pueden hacerlo, mientras que tenemos los suficientes motivos para creer que otros son totalmente insensibles a estas cuestiones.»

El tipo de investigación que sugieren estos ejercicios nos permitiría llegar bastante lejos en una explicación de la estructura del mundo perceptivo y comportamental de los murciélagos, de modo que podríamos ordenar las narraciones heterofenomenológicas en función de su realismo, descartando aquellas que atribuyeran o presupusieran talentos discriminativos o disposiciones reactivas que se puede demostrar que no forman parte de la ecología ni de la neurofisiología del murciélago. Aprenderíamos, por ejemplo, que los murciélagos no se ven perturbados por los chillidos que emiten para producir sus ecos, porque poseen un músculo que les tapa los oídos en perfecta sincronía con la emisión de los chillidos, no muy distinto de los dispositivos que evitan que los sistemas de radar más sensibles se vean perturbados por las señales que emiten. Se han llevado a cabo numerosas investigaciones en este terreno, y podemos decir mucho más, por ejemplo sobre por qué los murciélagos emiten chillidos con diferentes frecuencias, en función de si están buscando una presa, aproximándose a su objetivo, o están a punto de capturarla (Akins, 1989, 1990).

Cuando hemos conseguido elaborar un relato heterofenomenológico que ningún crítico será capaz de refutar, deberemos aceptarlo —provisionalmente, a la espera de que se realicen nuevos descubrimientos— como relaciones precisas sobre lo que se siente al ser la criatura en cuestión. Así es, después de todo, como nos tratamos los unos a los otros. Al recomendar que tratemos del mismo modo a los murciélagos y a otros candidatos para la interpretación, no estoy *trasladando* el peso de la prueba, sino ampliando el peso de la prueba normal, humana, a otras entidades.

Podríamos utilizar estas investigaciones para desmitificar todo tipo de ilusiones románticas sobre la conciencia de los murciélagos. Sabemos que el delicioso libro para niños de Randall Jarrell, *The Bat-Poet* (1963) ["El murciélago poeta"], es una fantasía, porque sabemos que los murciélagos no hablan. Algunas afirmaciones fantásticas nada evidentes sobre su fenomenología sucumben ante hechos aún menos evidentes, pero conocidos, sobre su fisiología y su conducta. Estas investigaciones nos enseñarían muchas cosas sobre de qué puede y de qué no puede ser consciente un murciélago bajo condiciones diversas, mostrándonos con qué recursos cuenta su sistema nervioso para representar esto o aquello, y verificando experimentalmente si el murciélago utilizó la información para modular su conducta. Resulta difícil imaginar la abundante información que podemos acumular con estas investigaciones hasta que uno ha empezado a llevarlas a cabo. (Para una investigación preliminar sorprendentemente detallada sobre lo que se siente al ser un mono verde, por ejemplo, véase Cheney y Seyfarth, *How Monkeys See the World,* 1990.)

Llegados a este punto, es lícito plantear la siguiente objeción: estas ob-

servaciones nos enseñarían mucho sobre la organización cerebral y el procesamiento de la información en el murciélago, lo cual nos mostraría simplemente *de qué no son conscientes* los murciélagos, dejando completamente abierta la cuestión sobre de qué *son* conscientes los murciélagos, si es que son conscientes de algo. Como es bien sabido, una gran parte del procesamiento de la información en los sistemas nerviosos es completamente inconsciente, de modo que estos métodos de investigación no nos permitirían rechazar la hipótesis de que los murciélagos son... zombíes voladores, ¡y de que no se sentiría nada al ser criaturas como éstas! (Wilkes, 1988, pág. 224, se pregunta qué ocurriría si la ecolocalización de los murciélagos fuera una especie de visión ciega, que no produce ningún efecto subjetivo.)

Ay, que el murciélago está tirando de la manta. Éste parece ser el mal cariz que está tomando nuestra discusión, y debemos intentar corregirlo. Richard Dawkins (1986), en un clarificador examen del diseño de la ecolocalización en los murciélagos de herradura *(Rhinolophidae)*, nos presenta una clara versión de la imagen que está empezando a surgir.

> Los sistemas de radar de la policía para detectar excesos de velocidad utilizan el efecto Doppler. (...) Comparando la frecuencia de salida con la frecuencia del eco de retorno, la policía, o, *mejor dicho, su instrumental automático* [la cursiva es mía], puede calcular la velocidad de los coches. (...) Comparando el tono de su chillido con el tono del eco de retorno, por tanto, el murciélago (o, *mejor dicho, su ordenador de abordo en el cerebro*) [la cursiva es mía] podría, en teoría, calcular a qué velocidad está volando hacia el árbol (pág. 30-31).

Resulta tentador preguntarse: ¿hay algo en el murciélago que esté situado en relación a su «ordenador de abordo» (el cual opera sin una brizna de conciencia) del mismo modo que los policías están situados en relación a su «dispositivo automático»? Los policías no tienen que calcular el efecto Doppler conscientemente, pero tienen que experimentar, conscientemente, la lectura de su dispositivo que dice, en brillantes símbolos de LED: «150 km/h». Éste es el dato que les hace saltar sobre sus motocicletas y poner en marcha sus sirenas. Es plausible suponer, por tanto, que el murciélago tampoco tiene que calcular conscientemente el efecto Doppler —su ordenador de abordo se ocupa de ello—, pero, ¿no queda nada en el murciélago que pueda asumir el papel del policía, un testigo capaz de apreciar (conscientemente) la «información de salida» del ordenador de análisis del efecto Doppler del murciélago? Nótese que podríamos sustituir fácilmente los agentes de policía por un dispositivo automático que registrara la matrícula del vehículo infractor, comprobara el nombre y dirección del propietario y le enviara la multa. No hay nada de especial en la tarea que llevan a cabo los policías que indique que no pudiera efectuarse sin que algo fuese experimentado. El mismo argumento vale, aparentemente, para el murciélago. Un murciélago podría ser un zombie. Sería un zombie —o esto es, cuando menos, lo que esta línea de argumentación sugiere— a menos que hubiese un

observador interno que reaccionase a la presentación interna de un modo parecido a como reaccionan los agentes a las luces rojas de sus instrumentos.

No caigamos en la trampa. Ésta es nuestra vieja némesis, la audiencia del Teatro Cartesiano. *Nuestra* conciencia no depende del hecho de que nuestro cerebro esté habitado por un agente interno para quien el cerebro prepara presentaciones, así que nuestra incapacidad de hallar un agente central en el cerebro del murciélago no altera nuestra atribución de conciencia, o nuestra afirmación de que somos capaces de decir cómo es su conciencia. Para comprender la conciencia de un murciélago, simplemente tenemos que aplicar al murciélago los mismos principios que nos aplicamos a nosotros mismos.

¿Pero qué podría hacer entonces un murciélago, que fuese lo bastante especial como para convencernos de que estamos en presencia de una conciencia genuina? Parece que, por muchos utilizadores de información de entrada que situemos ante el traductor Doppler del murciélago, nunca habría una razón de la «tercera persona», desde el exterior, lo bastante convincente como para atribuir experiencia consciente al murciélago. No es cierto. Si el murciélago pudiese hablar, por ejemplo, generaría un texto a partir del cual podríamos originar un mundo heterofenomenológico, lo cual nos proporcionaría el mismo fundamento para atribuirle conciencia a un murciélago que a una persona. Pero, como ya hemos dicho, los murciélagos no hablan. Sí pueden, sin embargo, mostrar ciertas conductas no verbales, las cuales bastan para sentar unas bases claras para describir su mundo heterofenomenológico o, como lo denominó von Uexküll (1909), el pionero en estas investigaciones, su *Umwelt und Innenwelt,* su mundo del entorno y su mundo interior.

La heterofenomenología sin texto no es imposible, sólo más difícil (Dennett, 1988a, 1988b, 1989a, 1989b). Una rama de la heterofenomenología animal es conocida con el nombre de etología cognitiva, el intento de modelar las mentes animales estudiando su conducta sobre el terreno y experimentando con ella. Hallamos una buena muestra de las posibilidades y dificultades con que se encuentra este tipo de investigación en Cheney y Seyfarth (1990), Whiten y Byrne (1988), y en Ristau (1991), un volumen de homenaje dedicado a Donald Griffin, el pionero en la investigación de la ecolocalización de los murciélagos y el creador del campo de la etología cognitiva. Una de las más frustantes dificultades con que se encuentran estos investigadores es que muchos de los experimentos que se le podrían ocurrir a uno resultan ser totalmente inútiles ante la ausencia de lenguaje; no se puede *preparar* a los sujetos (y saber que han sido preparados) sin conversar con ellos, como sería necesario para llevar a cabo uno de extos experimentos (Dennett, 1988a).

Éste no es solamente un problema epistemológico para el heterofenomenólogo; la dificultad misma de crear las circunstancias experimentales necesarias en el entorno natural demuestra algo mucho más fundamental sobre la mente de las criaturas sin lenguaje. Demuestra que las situaciones

ecológicas de estos animales nunca les han proporcionado las oportunidades para el desarrollo (por evolución, por aprendizaje, o ambos) de muchas de las actividades mentales avanzadas que conforman nuestras mentes, de modo que podemos estar completamente seguros de que nunca llegaron a desarrollarlas. Por ejemplo, considérese el concepto de *secreto*. Un secreto no es solamente algo que uno sabe y que los demás no saben. Para que uno pueda tener un secreto es preciso que sepa que los demás no lo saben y debe ser capaz de controlar este hecho. (Si usted es el primero que ve que se aproxima una estampida, puede que sepa algo que los demás no saben, pero no por mucho tiempo; usted no puede *guardarse* ese poquito de información secreta.) La ecología comportamental de una especie tiene que estar estructurada de forma especial para que los secretos puedan jugar algún papel. Los antílopes, en sus manadas, no tienen secretos ni posibilidades de tenerlos. Por tanto, un antílope es tan incapaz de trazar un plan secreto como de contar hasta cien o de disfrutar de los colores de una puesta de sol. Los murciélagos, que se embarcan en incursiones relativamente solitarias en las que podrían ser capaces de reconocer su aislamiento de sus rivales, cumplen una de las condiciones necesarias para tener secretos. ¿Tienen también intereses a los que la explotación de esos secretos pudiera ser de alguna utilidad? (¿Qué podría hacer una almeja con un secreto? ¿Quedarse ahí, en la arena, riéndose sola?) ¿Tienen los murciélagos también hábitos de sigilosidad y engaño durante la caza que pudieran adaptarse para actividades de mantenimiento de secretos más elaborados? De hecho, hay muchas preguntas de este tipo que, una vez planteadas, sugieren nuevas vías de investigación y de experimentación. La estructura mental de un murciélago es tan accesible como la estructura de su sistema digestivo; el modo de investigar el uno o el otro consiste en pasar sistemáticamente de un análisis de su contenido a un análisis del mundo a partir del cual se deriva ese contenido, prestando atención a los métodos y los objetivos de la derivación.

Wittgenstein escribió una vez: «Si un león pudiera hablar, no podríamos entenderlo» (1953, pág. 223).* Yo pienso, por el contrario que, si un león pudiera hablar, ese león tendría una mente tan diferente de la de los demás leones, que, aunque pudiéramos comprenderlo perfectamente, de él aprenderíamos muy poco sobre los leones normales. El lenguaje, como vimos en los capítulos anteriores, juega un papel fundamental en la estructuración de la mente humana, y no puede considerarse que la mente de una criatura sin lenguaje —y que, además, no necesita el lenguaje— esté estructurada de esa manera. ¿Significa esto que los animales sin lenguaje «no son conscientes» (como insistía Descartes)? Llegados a este punto, siempre surge esta pregunta como una especie de desafío de los incrédulos, pero no debemos sentirnos obligados a responderla tal como está planteada. Nótese que presupone una idea cuya influencia hemos intentado evitar: el supuesto de que la conciencia es una propiedad especial de todo o nada que divide el universo en

* Páginas 510/511 de la edición bilingüe (Crítica, 1988).

dos categorías totalmente distintas: las cosas que la tienen (se siente algo al ser esas cosas, como diría Nagel) y las cosas que no la tienen. Ni siquiera en nuestro caso podemos trazar una línea que separe nuestros estados mentales conscientes de nuestros estados mentales inconscientes. La teoría de la conciencia que hemos esbozado permite un margen de variación en cuanto a la arquitectura funcional, y, pese a que la presencia del lenguaje conlleva un notable crecimiento del poder imaginativo, la versatilidad, el autocontrol (por mencionar sólo algunos de los poderes más evidentes de la máquina virtual joyceana), estos poderes no tienen el poder *adicional* de encender una luz interna especial que, en caso contrario, estaría apagada.

Cuando nos imaginamos lo que se siente al ser una criatura sin lenguaje empezamos, naturalmente, a partir de nuestra propia experiencia, de modo que la mayor parte de las cosas que nos vienen a la mente deben seguir un proceso de reajuste (generalmente hacia menos). El tipo de conciencia que poseen estos animales, comparada con la nuestra, está radicalmente impedida. Un murciélago, por ejemplo, no sólo no puede preguntarse si hoy es viernes; ni siquiera puede preguntarse si es un murciélago; el acto de preguntarse no juega ningún papel en su estructura cognitiva. Aunque un murciélago, e incluso la simple langosta, tiene un yo biológico, no tiene un yo egoico del que hablar; no tiene un centro de gravedad narrativa o, si lo tiene, no es lo bastante relevante. No hay palabras en la punta de la lengua, pero tampoco lamentaciones, ni complejos anhelos, ni recuerdos nostálgicos, ni grandes planes, ni reflexiones sobre lo que se siente al ser un gato, ni siquiera sobre lo que se siente al ser un murciélago. Esta lista de rechazos no sería más que una muestra de escepticismo barato si no poseyéramos una teoría empírica positiva sobre la que basarlos. ¿Estoy afirmando haber probado que los murciélagos no podrían poseer estos estados mentales? No, pero tampoco puedo probar que las setas *no podrían* ser naves intergalácticas que nos espían.

¿No es acaso un terrible prejuicio antropocéntrico? Además, ¿qué decir de los sordomudos? ¿Son conscientes? Por supuesto que lo son; pero no saquemos conclusiones precipitadas sobre su conciencia, fruto de una simpatía mal dirigida. Cuando un sordomudo adquiere el lenguaje (el lenguaje de signos, que es el lenguaje más natural que puede aprender un sordomudo), nace una mente humana completa, claramente distinta de la mente de una persona que pueda oír, pero con la misma capacidad de producir complejas reflexiones y con el mismo poder generativo, incluso más. Pero sin un lenguaje natural, la mente de un sordomudo queda terriblemente empobrecida. (Véase Sacks, 1989, especialmente su bibliografía comentada.) Como señala el filósofo Ian Hacking (1990) en su reseña del libro de Sacks, «se necesita una imaginación muy viva para hacerse una idea de aquello de que carece un niño sordo». No hacemos ningún favor a los sordomudos al imaginar que, en la ausencia de lenguaje, disfrutan de todas las delicias mentales de que disfrutamos nosotros los humanos que podemos oír, ni se lo hacemos a los

animales no humanos al intentar ocultar los hechos patentes sobre las limitaciones de sus mentes.

Y todo esto, como muchos de ustedes están impacientes por señalar, es un subtexto que ha estado luchando desde hace tiempo por salir a la superficie: muchas personas temen una explicación de la conciencia porque temen que, si llegamos a explicarla, perderemos nuestros valores morales. Quizá podamos imaginar un robot consciente (o la conciencia de un murciélago), pero *no deberíamos intentarlo,* piensan. Si adoptamos esa mala costumbre, empezaremos a tratar a los animales como si fueran juguetes de cuerda, a los bebés y a los sordomudos como si fueran ositos de peluche y —llegando al colmo del insulto y la injuria— a los robots como si fueran personas de verdad.

3. Preocupación e importancia

Tomo prestado el título de este apartado de un artículo de Marian Stamp Dawkins (1987), quien ha llevado a cabo cuidadosas investigaciones sobre las implicaciones morales de la heterofenomenología animal. (Sus primeros trabajos aparecen reseñados en su libro *Animal Suffering: The Science o Animal Welfare,* 1980.) Como observa Dawkins, nuestras actitudes morales hacia los animales están plagadas de incoherencias.

> No tenemos más que pensar en diversos animales para mostrar nuestras incoherencias. Se organizan manifestaciones en contra de las matanzas de bebés foca, pero no se organizan en contra de las matanzas de ratas. Muchas personas que no tienen ningún problema en comer cerdo o cordero se horrorizan ante la idea de comer perro o caballo (pág. 150).

Dawkins señala que hay dos perspectivas distintas dentro de todo este embrollo: la capacidad de razonamiento y la capacidad de sufrimiento. Descartes insistió mucho en la incapacidad de razonar (cuando menos, de razonar del mismo modo que lo hacen los humanos) de los animales no humanos, lo cual provocó una famosa respuesta por parte del filósofo utilitarista británico Jeremy Bentham: «Un perro o un caballo adulto son, sin punto de comparación, animales mucho más racionales y con los que es más fácil conversar que un bebé de un día, una semana o incluso un mes. Pero supongamos que fuese al contrario, ¿qué importancia tendría? La pregunta no es si pueden *razonar,* ni tampoco si pueden *hablar,* sino ¿pueden *sufrir*?» (Bentham, 1789). Aparentemente, éstas son posturas morales opuestas, pero según argumenta Dawkins, «otorgar un valor ético a la capacidad de sufrimiento nos llevará eventualmente a valorar los animales que son inteligentes. Aun cuando empezáramos rechazando el criterio del razonamiento propuesto por Descartes, lo más probable es que sean los animales con capacidad de razonamiento los que posean una capacidad de sufrimiento» (pág. 153).

Los motivos de esta idea están implícitos en la teoría de la conciencia que hemos desarrollado. El sufrimiento no consiste en verse atacado por un estado inefable intrínsecamente malo, sino en ver arruinados las propias esperanzas, planes y proyectos de vida por circunstancias impuestas sobre nuestros deseos, frustrando nuestras intenciones, sean éstas cuales sean. La idea de que el sufrimiento es algo explicable por la presencia de una determinada propiedad intrínseca —el horror, por ejemplo— tiene tan poco fundamento como la idea de explicar la diversión por la presencia de una hilaridad intrínseca. Así pues, la presunta inaccesibilidad, el desconocimiento último, del sufrimiento del otro es algo tan engañoso como las fantasías sobre los qualia intrínsecos que desenmascaramos anteriormente, aunque mucho más pernicioso. Se deduce, por tanto —lo cual enciende una luz en nuestra intuición— que la capacidad de sufrimiento está en función de la capacidad de poseer unos sistemas articulados de deseos, expectativas y otros estados mentales complejos a largo plazo y con un alto poder discriminativo.

Los seres humanos no son las únicas criaturas con la inteligencia suficiente como para sufrir; el perro y el caballo de Bentham muestran con su comportamiento que poseen la suficiente complejidad mental como para distingir, e interesarse por, un amplio espectro de dolores y otras imposiciones que no son negligibles, aun cuando el abanico de posibilidades es menor comparado con las posibilidades del sufrimiento humano. Otros mamíferos, particularmente los simios, los elefantes y los delfines, parecen poseer una gama mucho más amplia.

Como compensación por tener que soportar tanto sufrimiento, las criaturas inteligentes son también las que más se divierten. Se necesita una economía cognitiva con un presupuesto suficiente para la exploración y la autoestimulación a fin de definir el espacio de deseos derivados recursivamente que hace posible la diversión. Se habrá dado el primer paso cuando la arquitectura disponible permita apreciar el significado de «¡para ya, que me encanta!». Versiones superficiales de este poder constructivo se manifiestan en algunas especies superiores, pero se necesita una imaginación exuberante, y un cierto tiempo de ocio —algo que la mayoría de las especies no se puede permitir—, para desarrollar una amplia gama de placeres. Cuanto mayor sea el alcance, mayor será el detalle; cuanto mayor sea el poder discriminativo de los deseos, peor será cuando éstos se vean frustrados.

Pero, ¿qué importancia puede tener, se preguntará usted, que los deseos de una criatura se vean frustrados si no son deseos *conscientes*? Mi respuesta: ¿por qué debería ser más importante si fueran conscientes, especialmente si la conciencia fuese, como muchos piensan, una propiedad que siempre escapará a todo intento de investigarla? ¿Por qué deberían ser las esperanzas frustradas de un «zombie» menos importantes que las esperanzas frustradas de una persona consciente? Estamos ante un juego de espejos que es preciso descubrir y desmantelar. La conciencia, nos dicen, es lo que importa, pero después se aferran a doctrinas de la conciencia que nos impiden sistemáticamente llegar a comprender *por qué* es importante. Postular la exis-

tencia de unas cualidades interiores especiales que no sólo son privadas e intrínsecamente valorables, sino también inconfirmables e investigables no es más que puro oscurantismo.

Dawkins muestra de qué modo es posible explorar experimentalmente las diferencias investigables —las únicas diferencias que podrían tener alguna importancia—, y merece la pena comprobar hasta dónde podemos llegar incluso con experimentos muy simples con especies muy poco atractivas en principio.

> Las gallinas criadas en libertad o en grandes corrales pasan mucho tiempo rascando el suelo, de modo que tuve la sospecha de que la falta de paja o arena en el suelo de las jaulas les podría causar algún sufrimiento. Sin dudarlo, cuando les ofrecí la oportunidad de escoger entre una jaula con el suelo de alambre y una con el suelo de paja y arena que les permitiera rascar, escogían la segunda jaula. De hecho, estaban dispuestas a entrar en una jaula diminuta (tan pequeña que apenas podían darse la vuelta) si así tenían acceso a un suelo de paja y arena. Incluso gallinas que siempre habían vivido en jaulas de alambre y que nunca habían tenido contacto con la paja y la arena escogían una jaula con suelo de paja y arena. Aunque estos resultados resultaban sugerentes, no había suficiente. Tenía que demostrar no sólo que las gallinas mostraban una preferencia por la paja y la arena, sino también que su preferencia era tan fuerte que nos permitiera afirmar que sufrían si se las privaba de ella.
>
> Se ofreció entonces a las gallinas una elección ligeramente distinta. En esta ocasión tenían que escoger entre una jaula con suelo de alambre que contenía comida y agua y una jaula con suelo de paja y arena sin alimento ni agua. (...) El resultado fue que pasaban mucho más tiempo en la jaula con suelo de paja y arena, y mucho menos en la jaula con suelo de alambre, aunque éste era el único lugar donde podían comer y beber. A continuación se introdujo una nueva complicación. Las gallinas tenían que «trabajar» para pasar de una jaula a otra. Tenían que saltar por un corredor o empujar una cortina de plástico negro. Ahora, pasar de una jaula a otra tenía un coste. (...) Las gallinas seguían pasando el mismo tiempo en la jaula con suelo de alambre y con comida que antes cuando no había ninguna dificultad para acceder a ella. Sin embargo, apenas accedían a la jaula con suelo de paja y arena. Simplemente, no parecían estar preparadas para trabajar o para pagar un coste a fin de poder entrar en la jaula. (...) Al contrario de lo que yo esperaba, las gallinas parecían estar diciendo que, en realidad, la paja y la arena no les importaban (págs. 157-159).

Dawkins concluye que «el sufrimiento por parte de una mente se muestra en animales que poseen una mente lo bastante racional como para hacer algo en cuanto a las condiciones que les hacen sufrir», y continúa señalando que «es también probable que los organismos sin la capacidad de hacer nada para apartarse de una fuente de peligro nunca desarrollarían la capacidad de sufrir. No habría ningún sentido evolutivo en un árbol que, mientras le van cortando las ramas, tuviese la capacidad de sufrir en silencio» (pág. 159). Como vimos en el capítulo 7 (véase también el capítulo 3, nota 9),

es preciso ser muy cuidadoso al construir argumentos evolutivos sobre la función como éstos. Sin embargo, en ausencia de datos positivos que avalen la imputación de sufrimiento, o de datos positivos que avalen la sospecha de que, por un motivo u otro, dichos datos permanecen ocultos, deberemos concluir que no hay tal sufrimiento. No debemos temer que esta regla tan austera nos lleva a faltar a nuestras obligaciones para con las demás criaturas. Nos proporcionan unas amplias bases para alcanzar conclusiones positivas: muchos animales, pero no todos, son capaces de mostrar una amplia gama de sufrimientos. Reconociendo estas diferencias de grado, podrá organizarse una defensa más persuasiva en favor de un tratamiento más humanitario, que promulgando caritativamente el insostenible dogma de la universalidad y la igualdad del sufrimiento animal.

Puede que esto resuelva el problema objetivo sobre la presencia o ausencia de sufrimiento, pero no resuelve el problema de los sentimientos morales afectados por la perspectiva de ver la conciencia explicada de manera tan descorazonadoramente mecanicista. Todavía hay mucho en litigio.

Poseo una granja en Maine, y estoy encantado de que haya osos y coyotes viviendo en mis bosques. Casi nunca los veo, y apenas percibo indicios de su presencia, pero me gusta saber que están ahí, y me sentiría muy infeliz si me enterara de que se han marchado. Tampoco me sentiría completamente compensado por la pérdida si uno de mis amigos investigadores en IA me llenara los bosques de animales robot (aunque, bien pensado, la idea no deja de tener su encanto). Me importa que haya criaturas salvajes que descienden de criaturas salvajes, y que viven cerca de mí. Igualmente, me encanta que haya muchos conciertos en el área de Boston que *ni siquiera* escucho e, incluso, de cuya existencia *ni siquiera* llego a enterarme.

Estos hechos son de una naturaleza especial. Son hechos que nos parecen importantes simplemente porque una parte del entorno que nos importa es nuestro entorno de creencias. Y puesto que no es fácil que sigamos creyendo en proposiciones cuya evidencia se ha evaporado, nos importa que las creencias sean *ciertas,* aun cuando no dispongamos de ninguna evidencia directa que las confirme. Como cualquier parte del entorno, un entorno de creencias puede ser frágil, compuesto de partes que están interconectadas tanto por accidentes históricos como por vínculos bien diseñados. Considérese, por ejemplo, la delicada parte de nuestro entorno de creencias relacionado con la suerte de nuestros cuerpos después de la muerte. Muy pocos creen que el alma permanezca en el cuerpo después de la muerte, ni siquiera los que creen en el alma creen en algo así. Y sin embargo, muy pocos de nosotros toleraría una «reforma» que animara a la gente a envolver sus muertos en plástico y a tirarlos a la basura, o a utilizar cualquier otro procedimiento tan poco ceremonioso como éste para deshacerse de ellos. ¿Por qué no? Sin duda no es porque creamos que los cadáveres puedan sufrir por un ataque a su dignidad. Un cadáver puede sufrir por su dignidad tanto como un tronco. Y sin embargo la idea es chocante, repulsiva. ¿Por qué?

Los motivos son complejos, pero podemos establecer algunos simples pun-

tos. Una persona no es solamente un cuerpo; una persona *tiene* un cuerpo. Ese cadáver es el cuerpo del viejo y querido Jones, un centro de gravedad narrativa que debe su realidad tanto a nuestros esfuerzos conjuntos de mutua interpretación heterofenomenológica como a ese cuerpo que ahora yace ahí sin vida. Los límites de Jones no son idénticos a los límites del cuerpo de Jones, y los intereses de Jones, gracias a esa curiosa práctica humana de tejer un yo, pueden extenderse más allá de los intereses biólogicos básicos que promovieron esa práctica. Tratamos a su cadáver con respeto porque es importante para la conservación del entorno de creencias en que todos nosotros vivimos. Si empezamos a tratar a los cadáveres como basura, por ejemplo, también podría cambiar la manera en que empecemos a tratar a los casicadáveres: aquellos que aún viven, pero están muriendo. Si no pécaramos por exceso en la prolongación de los rituales y las prácticas de respeto más allá del umbral de la muerte, el que muere (y los que se preocupan por él) se enfrentarían a una angustia, una afrenta, una *posibilidad,* que corre el riesgo de ofenderlos. Puede que tratar «mal» a un cadáver no dañe directamente a la persona muerta, y, sin duda, no daña al cadáver, pero si se convirtiera en práctica común y ello llegara a saberse (como así sería), se produciría un cambio significativo del entorno de creencias que rodea a la muerte. Las personas imaginarían los acontecimientos que seguirían a su fallecimiento de forma totalmente distinta a como los imaginamos ahora, lo cual sería bastante deprimente. Quizá sin motivo, pero, ¿qué más da? Si las personas van a estar deprimidas, ése es motivo suficiente para adoptar una determinada política.

Así pues, existen razones de peso indirectas, pero aún estimables y legítimas, para seguir respetando a los cadáveres. No necesitamos ninguna mitología sobre algo especial que reside en los cadáveres que los convierte en entidades privilegiadas. *Podría* ser un mito útil para extenderlo entre los más ignorantes, pero sería de un paternalismo extremo el pensar que los que estamos mejor informados debemos conservar dichos mitos. Similarmente, existen muy buenos motivos para tratar a todos los animales con cuidado y con atención. Estos motivos son, en parte, independientes del hecho de que determinados animales sientan ciertos tipos de dolor. Dependen de modo más directo del hecho de que diversas creencias formen parte de nuestra cultura, y nos importan, tanto si *deben* importarnos como si no. Puesto que ahora son importantes, son importantes. Sin embargo, la racionalidad del entorno de creencias —el hecho de que las creencias estúpidas o sin fundamento tiendan a desaparecer a largo plazo, a pesar de la superstición— implica que las cosas que ahora importan no tienen por qué importar siempre.

Pero entonces, como anticipábamos en el capítulo 2, una teoría que ataque directamente al entorno de creencias general posee la capacidad de hacer un verdadero daño, de provocar sufrimiento (en las personas con una particular preocupación por los animales, por ejemplo, cualquier cosa que les ocurra a los animales es sufrimiento). ¿Significa esto que debemos abandonar todo intento de investigar estos asuntos por temor a abrir la caja de

Pandora? Ello podría tener alguna justificación si pudiéramos convencernos de que nuestro entorno de creencias, basado en mitos o no, es, sin lugar a dudas, moralmente aceptable, pero me permito afirmar que está claro que no es así. Aquellos que están preocupados por los costes con que nos amenaza está ilustración que nadie ha pedido deberían tomarse la molestia de analizar los costes de los mitos actuales. ¿Realmente pensamos que aquello con lo que actualmente nos enfrentamos merece ser protegido con una especie de oscurantismo creativo? ¿Pensamos, por ejemplo, que es preciso dedicar recursos para conservar las imaginarias perspectivas de una posible vida mental renovada de las personas en estado de coma profundo, mientras no hay recursos para mejorar la situación desesperada, y muy real, de los más pobres? Los mitos sobre la santidad de la vida, o de la conciencia, son un arma de doble filo. Puede que sean útiles para levantar barreras (contra la eutanasia, contra la pena de muerte, contra el aborto, contra el comer carne) a fin de impresionar a los que no tienen imaginación, pero al precio de provocar una hipocresía ofensiva o un autoengaño ridículo entre los más ilustrados.

Las barreras absolutistas, como la línea Maginot, rara vez cumplen la función para la que fueron ideadas. La campaña que se organizó en contra del materialismo ya ha sucumbido ante su propio desconcierto, y la campaña en contra de la «IA fuerte», aunque igual de bienintencionada, sólo puede ofrecernos los más gastados modelos alternativos de la mente. Sin duda sería mejor intentar fomentar la estima por unas bases *no* absolutistas, *no* intrínsecas y *no* dicotomizadas de nuestras preocupaciones morales que pudiesen coexistir con nuestro creciente conocimiento del funcionamiento interno de la más fascinante de las máquinas, el cerebro. Los argumentos morales, *esgrimidos por ambos bandos*, sobre asuntos como la pena de muerte, el aborto, el comer carne y la experimentación con animales no humanos, por ejemplo, se situarán a un más alto nivel cuando rechacemos los mitos que no merecen ninguna protección.

4. La conciencia explicada, ¿o eliminada?

Cuando aprendemos que la única diferencia entre el oro y la plata es el número de partículas subatómicas en sus átomos, puede que nos sintamos estafados o enfadados; esos físicos han eliminado algo: le han quitado el carácter dorado al oro; han suprimido el carácter plateado de la plata que tanto apreciamos. Y cuando explican de qué modo la reflexión y la absorción de las radiaciones electromagnéticas dan cuenta de los colores y la visión en color, parecen estar olvidando lo que es más importante. Pero si no «suprimiéramos» algo, no podríamos empezar a explicar. El hecho de eliminar algo no es un rasgo de las explicaciones fallidas, sino de las explicaciones logradas.

Sólo una teoría que explicara acontecimientos conscientes en términos

de acontecimientos inconscientes podría aspirar a explicar la conciencia. Si su modelo de cómo el dolor es un producto de las actividades cerebrales todavía tiene una cajita con la etiqueta «dolor», usted ni siquiera ha empezado a explicar lo que es el dolor, y si su modelo de la conciencia funciona perfectamente hasta el mágico instante en que tiene que decir «y entonces se produce un milagro», es que usted ni siquiera ha empezado a explicar lo que es la conciencia.

Esto lleva a algunas personas a insistir en que la conciencia nunca podrá ser explicada. Pero, ¿por qué la conciencia debería ser la única cosa que no se puede explicar? Los sólidos, los líquidos y los gases pueden explicarse a partir de cosas que no son ni sólidos ni líquidos ni gases. Sin duda la vida puede explicarse a partir de cosas que no están vivas; y esa explicación no deja sin vida a las cosas vivas. La ilusión de que la conciencia es la excepción proviene, sospecho, del hecho generalizado de no haber sabido comprender este rasgo general de la explicación. Al pensar, sin razón, que la explicación comporta suprimir algo, creemos salvar aquello que en caso contrario se perdería, volviéndolo a colocar en el observador como un quale, o como alguna propiedad «intrínsecamente» maravillosa. La psique se convierte en el manto protector tras el cual se ocultan todos esos gatitos adorables. Puede que haya *motivos* para pensar que la conciencia no se puede explicar, pero, como espero haber demostrado, existen buenas *razones* para pensar que sí se puede.

Mi explicación de la conciencia dista mucho de ser completa. Podría decirse que no es más que el principio, pero es que *es* el principio, porque rompe el encantamiento del círculo mágico de ideas que hicieron que pareciera imposible la explicación de la conciencia. No puede decirse que haya sustituido una teoría metafórica, el Teatro Cartesiano, por una teoría *no* metafórica («literal, científica»). La verdad es que todo lo que yo he hecho no es más que sustituir una familia de metáforas e imágenes por otra, cambiando el teatro, el testigo, el Significador Central, el figmento, por el *software*, las máquinas virtuales, las Versiones Múltiples, el pandemónium de homúnculos. Así que no es más que una guerra de metáforas, me dirán ustedes, pero las metáforas no son «sólo» metáforas; las metáforas son herramientas de pensamiento. Nadie puede pensar sobre la conciencia sin ellas, de modo que es importante equiparse con el mejor juego de herramientas posible. Vean lo que hemos construido con nuestras herramientas. ¿Podrían haberlo imaginado sin ellas?

APÉNDICE A

(PARA FILÓSOFOS)

Hay momentos en el libro en que trato muy de soslayo, y sin apenas comentarios, algunas de las principales batallas filosóficas, o bien no cumplo con las obligaciones académicas propias de un filósofo. Aquellos filósofos que han leído el manuscrito del libro me han planteado algunas preguntas en relación a estas lagunas. Dichas preguntas se refieren a aspectos que pueden no ser de interés para los que no se dedican profesionalmente a la filosofía, pero merecen recibir una respuesta.

Al final del capítulo 11, en el diálogo con Otto, usted se da demasiada prisa en introducir los «presentimientos» en tanto que actos de habla sin actor y sin habla, y después revisar su propia caricatura, sustituyendo los presentimientos por «acontecimientos de fijación de contenido» sin ningún comentario adicional. ¿No es éste uno de los pasos clave de toda su teoría?

Efectivamente. Éste es el principal punto de contacto con la otra mitad de mi teoría de la mente, la teoría del contenido o de la intencionalidad cuyo desarrollo más reciente presenté en mi libro *La actitud intencional.* Hay muchos otros puntos en el libro donde me remito a esta teoría, pero creo que ustedes han sabido detectar el punto fundamental. Sin esa teoría del contenido, éste sería el lugar donde mi teoría diría «y, entonces, se produce un milagro». Mi estrategia fundamental siempre ha sido la misma: primero, desarrollar una explicación del contenido que sea *independiente de* y *más fundamental que,* la conciencia, una explicación del contenido que trate de manera uniforme toda fijación de contenido inconsciente (en los cerebros, en los ordenadores, en el «reconocimiento» por parte de la evolución de las propiedades de los diseños seleccionados); y, segundo, construir una explicación de la conciencia sobre estas bases. Primero el contenido, después la conciencia. Las dos mitades de *Brainstorms* recapitulaban esta estrategia, pero a medida que ambas mitades crecieron, cada una dio lugar a un volumen independiente. Este libro completa la tercera etapa de esta campaña. Evidentemente, esta estrategia es exactamente la opuesta a la visión de Nagel y Searle, quienes, cada uno a su manera, insisten en tratar la conciencia como el elemento fundamental. El motivo por el cual en el capítulo 11 pasé con tanta rapidez por este asunto tan importante es que no supe hallar una manera satisfactoria de sintetizar los cientos de páginas de análisis y argumentación que he dedicado a mi teoría del contenido en algo que fuera a la vez preciso y accesible. Así que, si ustedes consideran que fui demasiado depri-

sa en aquellas páginas, les ruego que consulten la versión más pausada en las páginas que se citan en la bibliografía.

Sin embargo, parece haber una tensión —si no una total contradicción— entre las dos mitades de la teoría. La actitud intencional presupone (o fomenta) la racionalidad, y, por tanto, la unidad, del agente —el sistema intencional— mientras que el modelo de las Versiones Múltiples se opone por completo a esa visión unitaria. ¿De acuerdo con su visión, cuál es, pues, la manera correcta de concebir la mente?

Todo depende de lo lejos que uno esté. Cuanto más nos acercamos, más importancia adquieren la falta de unidad, la multiplicidad y la competitividad. Después de todo, la base principal del mito del Teatro Cartesiano consiste en extrapolar, perezosamente, la actitud intencional a *todos los niveles interiores*. El tratar una entidad compleja y semoviente como un agente dotado de una única mente es una excelente manera de percibir una regularidad en sus actividades; dicha táctica nos resulta natural, y es muy posible que esté genéticamente favorecida en tanto que manera de percibir y de pensar. Pero cuando aspiramos a desarrollar una ciencia de la mente, debemos aprender a restringir y redirigir estos hábitos de pensamientos, subdividiendo el agente con una sola mente en una serie de miniagentes y microagentes (sin un único jefe). Llegados a este punto podemos ver que muchos de los fenómenos *aparentes* de la experiencia consciente han sido objeto de una descripción errónea por parte de la táctica unitaria tradicional. El sistema de amortiguar la tensión consiste en llevar a cabo esas forzadas identificaciones entre entidades heterofenomenológicas (según las concibe la perspectiva tradicional) y acontecimientos de fijación de contenido en el cerebro (tal como se los concibe en la nueva perspectiva).

Los filósofos a menudo han señalado las idealizaciones de la táctica tradicional, pero rara vez han llegado a un acuerdo en cuanto a ellas. Por ejemplo, empezando por Hintikka (1962), existe una voluminosa bibliografía dedicada a tratar las dificultades de la lógica de los estados reflexivos de creencia y conocimiento. Una de las idealizaciones esenciales en la formalización de Hintikka, como él mismo apunta, consiste en el hecho de que los enunciados regidos por la lógica que él desarrolla «deben ser proferidos en la misma y única ocasión. (...) La noción de olvido no es aplicable dentro de los límites de una ocasión» (pág. 7). La importancia de esta limitación, señala Hintikka, no siempre ha sido apreciada en su justa medida, y casi siempre ha acabado por perderse en la bruma de las controversias subsiguientes. Hintikka reconoce que esta cuantificación de «ocasiones» es una simplificación necesaria, básica para abordar, como él ha hecho, la formalización de los conceptos cotidianos de creencia y conocimiento; el contenido se fija en un instante y así se fija la identidad de la proposición en cuestión. En este libro he afirmado que dicha individuación artificial de «estados» y «tiempos» es uno de los rasgos que convierte en fantasías estos conceptos de la psicología del sentido común, cuando intentamos proyectarlos sobre las complejidades de lo que ocurre en el cerebro.

¿Finalmente, qué dice usted que son las experiencias conscientes? ¿Defiende usted la teoría de la identidad, es usted un materialista eliminativista, un funcionalista, un instrumentalista?

Me resisto a ceder a las presiones de los que me piden que defina una única proposición, formal, debidamente cuantificada, y que exprese el rasgo definitorio de mi teoría. Llenar los blancos en la fórmula *(x) (x es una experiencia consciente si, y sólo si...)* y defenderla en contra de los contraejemplos propuestos no es un buen método para desarrollar una teoría de la conciencia, y creo haber demostrado por qué. El carácter indirecto del método heterofenomenológico constituye precisamente el modo de evadir esas obligaciones inmotivadas de «identificar» o «reducir» las (presuntas) entidades que habitan en la ontología de los sujetos. *¿Identifican* los antropólogos a Fenhomo con el tipo que, según han descubierto, ha llevado a cabo todas esas buenas acciones en la jungla, o son «eliminativistas» con respecto a Fenhomo? Si han hecho bien *su* trabajo, el único asunto que queda por resolver puede zanjarse desde el punto de vista puramente diplomático, y no desde el punto de vista de la doctrina filosófica o científica. En cierto modo, podría decirse que mi teoría identifica experiencias conscientes con acontecimientos que llevan información, ya que eso es todo lo que ocurre, y muchos de los acontecimientos cerebrales tienen un sorprendente parecido con los habitantes de los mundos heterofenomenológicos de los sujetos. Sin embargo, otras propiedades de las entidades heterofenomenológicas —como por ejemplo, la posición que ocupan los elementos en la secuencia temporal subjetiva— podrían calificarse de «esenciales», en cuyo caso no *podrían* identificarse con los acontecimientos cerebrales disponibles, que pueden estar organizados en una secuencia distinta, sin correr el riesgo de violar la ley de Leibniz.

La cuestión de si es preciso tratar parte del mundo heterofenomenológico de un sujeto como una ficción útil, en lugar de tratarlo como una verdad un tanto forzada no es una cuestión que siempre haya recibido la atención que se merece. ¿Son reales las imágenes mentales? Existen estructuras de datos reales en el cerebro de las personas que son como imágenes, ¿son *éstas* las imágenes mentales por que preguntábamos? Si es que sí, entonces la respuesta es sí; si es que no, entonces la respuesta es no. ¿Son los qualia definibles funcionalmente? No, porque no existe ninguna propiedad como los qualia. O no, porque los qualia son propiedades disposicionales de los cerebros que no son estrictamente definibles en términos *funcionales*. O sí, porque si usted realmente lo comprendió todo sobre el funcionamiento del sistema nervioso, entonces usted lo comprende todo acerca de las propiedades de las que realmente están hablando las personas cuando afirman estar hablando sobre los qualia.

¿Soy, pues, un funcionalista? Sí y no. No soy un funcionalista de máquina de Turing, aunque dudo que nadie lo haya sido nunca, lo cual es una pena, porque, entonces, muchas refutaciones tendrán que ir a parar a la papelera. Soy una especie de «teleofuncionalista»; de hecho, quizá sea el primer te-

leofuncionalista (en *Content and Consciousness*), pero como espero haber dejado claro desde entonces, y he procurado subrayar aquí en mi discusión sobre la evolución y los qualia, no he cometido el error de intentar definir todas las diferencias mentales relevantes en términos de *funciones* biológicas. Ello comportaría una lectura totalmente equivocada de Darwin.

¿Soy un instrumentalista? Creo haber demostrado, en «Real patterns» (1991a), por qué ésta es una pregunta mal enfocada. ¿Son reales los dolores? Son tan reales como los cortes de pelo, los dólares, las oportunidades y las personas, y los centros de gravedad, pero ¿en qué medida son reales estas cosas? Estas preguntas con la voluntad de crear dicotomías surgen de la necesidad de rellenar el blanco en la fórmula cuantificada de más arriba, y algunos filósofos piensan que uno desarrolla una teoría de la mente construyendo una proposición antibalas como ésa y después defendiéndola. Una única proposición no es una teoría, es un eslogan; y lo que hacen algunos filósofos no es teorizar, es urdir eslóganes. ¿*Para* qué sirve esto? ¿Qué confusiones se disiparían, qué avances en cuanto al punto de vista se producirían, si alguien tuviera éxito en una empresa como ésta? ¿Realmente necesitamos algo que llevar impreso en la camiseta? Algunos urdidores de eslóganes son muy, muy buenos, pero como dijo en una ocasión memorable el psicólogo Donald Hebb, «si no merece la pena hacerlo, no merece la pena hacerlo bien».

Con esto no quiero decir que el ejercicio de efectuar definiciones precisas, y de criticar esas definiciones mediante contraejemplos, no tenga ningún valor. Considérese, por ejemplo, la definición del color. Los análisis más recientes y los intentos de definición llevados a cabo por los filósofos han aportado nueva luz al asunto. Han iluminado los conceptos y han eliminado malentendidos. Habida cuenta del cuidado que han puesto últimamente los filósofos en sus intentos de dar una definición precisa del color, mi rápida afirmación, en el capítulo 12, de que los colores son «propiedades reflectivas de los objetos o de los volúmenes transparentes» está demasiado poco justificada. ¿*Qué* propiedades reflectivas exactamente? Creo haber explicado por qué el intentar responder de forma precisa a esta pregunta sería una pérdida de tiempo; la única respuesta precisa no sería una respuesta concisa, por motivos que es muy fácil comprender. Ello significa que es muy difícil hallar una definición que «no sea circular». ¿Y qué? ¿Creo realmente que de este modo conseguiré enfrentarme a todos los problemas que han planteado mis contrincantes? (Además de los citados anteriormente, quisiera mencionar a Strawson, 1989, y Boghossian y Velleman, 1989, 1991.) Sí, pero es una larga historia, así que dejaré la pelota en su campo.

¿No es su posición, por tanto, más que una especie de verificacionismo?

Recientemente los filósofos han conseguido convencerse a sí mismos —y a más de un espectador inocente— de que el verificacionismo *siempre* es un pecado. Bajo la influencia de Searle y Putnam, por ejemplo, el investigador del cerebro Gerald Edelman se retracta precipitadamente de un acto de cuasiverificacionismo: «La falta de datos que evidencien la ausencia de auto-

conciencia en otros animales además de los chimpancés no nos permite considerar que no sean autoconscientes» (1989, pág. 280). ¡Venga! ¡Ánimo! Evidentemente, no podemos solamente *considerar* que no lo son, pero podemos investigar esa consideración, y si encontramos datos positivos suficientes para negarlo, entonces deberemos negarlo. Ha llegado el momento de que el péndulo produzca una nueva oscilación. En un comentario a unas críticas mías a Nagel (Dennett, 1982a), Richard Rorty escribió:

> Dennett piensa que uno puede mostrarse escéptico ante la insistencia de Nagel sobre la riqueza fenomenológica de la vida interior de los murciélagos «sin convertirse así en un verificacionista de pueblo». Yo no. Yo pienso que el escepticismo frente a intuiciones como las de Nagel o Searle es plausible sólo si se basa en consideraciones metodológicas generales sobre la naturaleza de esas intuiciones. La típica queja del verificacionista ante el realista es que el segundo insiste en fijarse en diferencias (entre, por ejemplo, murciélagos con y sin vida privada, perros con y sin intencionalidad intrínseca) que no tienen importancia; que sus intuiciones no pueden integrarse en un esquema explicativo porque son «[ruedas] que no pertenecen a la máquina» (Wittgenstein, 1953, I, 271). Me parece que ésta es una queja razonable, y la única queja que tenemos que expresar (Rorty, 1982a, págs. 342-343; véase también Rorty, 1982b).

Yo estuve de acuerdo, pero propuse una ligera corrección (0,742) de dicha afirmación: «Gracias a los vítores del profesor Rorty..., estoy dispuesto a presentarme como una especie de verificacionista, pero no, por favor, como un verificacionista de pueblo; seamos todos verificacionistas *urbanos*» (Dennett, 1982b, pág. 355). El presente libro sigue adelante en esta misma dirección, argumentando que si no nos convertimos todos en verificacionistas urbanos, acabaremos por tolerar cualquier cosa: el epifenomenismo, los zombíes, los espectros invertidos indistinguibles, los ositos de peluche conscientes y las arañas autoconscientes.

El punto más sobresaliente en favor del tipo de verificacionismo que defiendo lo encontramos en el capítulo 5, en la argumentación donde pretendo demostrar que, dado que no hay *ni puede haber* nada que favorezca los modelos orwellianos o estalinianos de la conciencia, entonces no hay nada que discutir al respecto. La refutación tradicional de este tipo de afirmaciones verificacionistas consiste en decir que se está prejuzgando el curso de la ciencia; ¿cómo sé yo que nuevos descubrimientos en neurociencia *no aportarán* nuevas bases para establecer tal distinción? La réplica —poco escuchada últimamente— es simple: sobre algunos conceptos (no todos, pero algunos) podemos estar seguros de que sabemos lo suficiente como para saber que *cualquier* novedad que provenga de la nueva ciencia nunca podrá reavivar esta posibilidad. Considérese, por ejemplo, la hipótesis de que el universo está cabeza arriba y su negación, la hipótesis de que está cabeza abajo. ¿Tienen sentido estas hipótesis? ¿Hay, o podría haber, algo que discutir? ¿Constituye acaso un pecado de verificacionismo el opinar que por muchas

revoluciones cosmológicas que se produzcan, ninguna convertirá *este* «debate» en un hecho que pueda ser resuelto empíricamente?

Pero entonces, usted es una especie de conductista, ¿no?

Esta pregunta ya se planteó con anterioridad, y me alegro de poder adherirme a la respuesta que dio Wittgenstein (1953).

> 307. «¿No eres después de todo un conductista enmascarado? ¿No dices realmente, en el fondo, que todo es ficción excepto la conducta humana?» —Si hablo de una ficción, se trata de una ficción *gramatical.*
>
> 308. ¿Cómo se llega al problema filosófico de los procesos y estados mentales y del conductismo? —El primer paso queda totalmente desapercibido. ¡Hablamos de procesos y estados y dejamos indeterminada su naturaleza! Quizás alguna vez lleguemos a saber más sobre ellos —pensamos. Pero justamente con ello nos hemos atado a un determinado modo de considerar las cosas. Pues tenemos un concepto definido de lo que quiere decir aprender a conocer más de cerca un proceso. (El paso decisivo en el truco del prestidigitador se ha dado y precisamente el que nos parecía inocente.) —Y ahora se desmorona la comparación que debía habernos hecho comprensibles nuestros pensamientos. Hemos de negar, pues, el proceso aún incomprendido en el medio aún inexplorado. Y así parece, por tanto, que hemos negado el proceso mental. ¡Y naturalmente no queremos negarlo!*

Algunos filósofos entienden que lo que yo estoy haciendo no es más que retomar los ataques de Wittgenstein en contra de los «objetos» de la experiencia consciente. Y de hecho así es. Como se expresa claramente en §308, si queremos evitar creernos el truco del prestidigitador, tenemos que comprender la «naturaleza» de los estados y procesos mentales *primero.* Este es el motivo por el cual necesité nueve capítulos hasta llegar al punto en que pude enfrentarme a los problemas con su disfraz filosófico, es decir, con su mal disfraz. Mi deuda con Wittgenstein es antigua y muy grande. Ya antes de licenciarme, él era mi héroe, así que me fui a Oxford, donde parecía ser el héroe de todo el mundo. Cuando comprobé que la mayoría de mis compañeros (a mi modo de ver) no estaban entendiendo nada, dejé de intentar «ser» wittgensteiniano, me limité a tomar lo que creía haber aprendido de las investigaciones y me propuse hacer que funcionara.

* De la edición bilingüe, Barcelona. Crítica, 1988.

Apéndice B

(PARA CIENTÍFICOS)

A menudo se acusa a los filósofos, con razón, de librarse a la psicología de café, (o la neurociencia o la física o...), y hay muchas historias un tanto embarazosas sobre filósofos cuyas vehementes declaraciones *a priori* se han visto refutadas en el laboratorio. Una actitud razonable para evitar este riesgo consiste en que el filósofo se retire sigilosamente hacia aquellos ámbitos conceptuales en que no hay peligro de cometer el error de decir algo que pueda verse refutado (o confirmado) por descubrimientos empíricos. Otra actitud razonable consiste en estudiar, aunque sea en el café, los mejores logros del laboratorio, los mejores esfuerzos de aquellos teóricos más preocupados por los aspectos experimentales y después seguir adelante con la filosofía, intentando iluminar los obstáculos conceptuales e, incluso, poniéndose a veces en situaciones comprometidas, con el objetivo de ser claro en cuanto a las implicaciones de ciertas ideas teóricas. Cuando se trata de asuntos conceptuales, los científicos no son menos inmunes a la confusión que los profanos. Después de todo, ellos también pasan mucho tiempo en el café, intentando hacerse una idea de cómo interpretar los resultados de los experimentos, y lo que hacen en estos momentos, se funde imperceptiblemente con lo que hacen los filósofos. Un asunto arriesgado, pero estimulante.

He aquí, pues, unas cuantas ideas no del todo maduradas sobre experimentos diseñados para evaluar las implicaciones del modelo de la conciencia que he esbozado, seleccionadas a partir de una serie de ideas todavía muy verdes que o bien no superaron el examen de mis pacientes informantes o bien se me demostró que ya habían sido puestas en práctica. (Mi buena media de ideas pertenecientes al segundo grupo me anima a seguir adelante.) Dado que, como filósofo, he intentado mantener mi modelo a un nivel tan general y tan poco comprometido como me ha sido posible, si hice bien mi trabajo estos experimentos deberían ayudar a determinar solamente *qué versión* de mi modelo se confirmaría; si el modelo se viera refutado por completo, me sentiría realmente desmentido e incómodo.

Sobre el tiempo y la organización temporal

Si las secuencias subjetivas son un producto de la interpretación, y no están directamente en función de las secuencias reales, debería ser posible

crear fuertes efectos interpretativos de diversos tipos que serían independientes de la organización temporal real.

1. *Andares de araña:* ligeros toques en secuencia, emulando el conejo cutáneo, pero orientados a producir juicios ilusorios sobre la *dirección.* Un caso típico consistiría en dos toques, separados en el espacio y en el tiempo por el mismo margen que en el fenómeno phi visual, donde la tarea consistiría en emitir un juicio sobre la dirección del «andar» (que es lógicamente equivalente a la *secuencia,* pero es un juicio más «inmediato», desde el punto de vista fenomenológico). Predicción: efectos típicos de fenómeno phi dependientes del ISI, con mayor agudeza en superficies de más alta resolución como la punta de los dedos o los labios.

Pídasele ahora al sujeto que mantenga sus dos dedos índices juntos y efectúese el primer toque en la punta de uno de ellos y el segundo en el otro. En este caso, debería producirse una resolución peor en cuanto a la direccionalidad, debido al requerimiento de que ahora las comparaciones tienen que ser bilaterales. Seguidamente, añádase una «ayuda» visual; permítase que el sujeto pueda ver la simulación con los dedos, pero proporciónesele información falsa: amáñese el aparato de modo que la *direccionalidad visual* sea la opuesta a la direccionalidad sugerida por la secuencia de toques. Predicción: los sujetos emitirán juicios falsos sin dudarlo, desautorizando o descartando la información sobre la secuencia real emitida por los receptores cutáneos. Si el efecto es muy fuerte, puede incluso hacer que se descarten juicios unilaterales o sobre el mismo dedo que, sin la información visual, serían muy precisos.

2. *Películas invertidas:* se pide a los sujetos que distingan breves «tomas» de cine o de vídeo, algunas de las cuales han sido invertidas, o en las cuales se han introducido pequeñas anomalías o rupturas en la secuencia. Los montadores de cine conocen muchos trucos, y acumulan un gran saber sobre los efectos de romper la secuencia de los fotogramas de una película. A veces montan deliberadamente escenas con los fotogramas desordenados a fin de crear efectos especiales: para aumentar la tensión o la impresión en las escenas de terror, por ejemplo. Algunos acontecimientos tienen una secuenciación natural muy fuerte; a todos nos ha divertido ver en una película a un bañista salir del agua por los pies para llegar saltando, ágil y seco, al trampolín. Otros acontecimientos son reversibles sin que apenas pueda ser percibido —una bandera ondeando, por ejemplo—, mientras que otros se sitúan en un punto intermedio; se necesitaría mucha atención para detectar si la película de una pelota botando está pasando hacia adelante o hacia atrás. Predicción: las personas no acertarán a distinguir inversiones en películas que carecen de sesgo interpretativo, en que la secuencia debe ser detectada y recordada. Por ejemplo, al mantener constantes las discrepancias en cuanto al movimiento, el tamaño y la forma, los sujetos deberían cometer errores al discriminar (reidentificar) secuencias que no invitan a una interpretación u otra en cuanto a la direccionalidad, y al distinguirlas de versiones inverti-

das o que incluyeran alguna otra transformación. (Los experimentos de discriminación de melodías serían el equivalente auditivo.)

3. *Escribiendo sobre su pie:* es un experimento diseñado para dar al traste con los juicios basados en la interpretación de los «tiempos de llegada» al «punto de central de disponibilidad». Supóngase que usted toma un lápiz y escribe unas letras sobre un costado de su pie desnudo, sin poder ver lo que está escribiendo. Las señales procedentes de los receptores cutáneos de su pie «confirmarían» que sus actos intencionados de escribir unas determinadas letras estaban siendo llevados a cabo por el lápiz que sostenía su mano. Añádase ahora visión indirecta, un monitor de televisión que mostrara su mano escribiendo sobre su pie, pero con la cámara colocada de tal manera que la punta del lápiz quedara oculta por la mano que lo sostiene. Estas señales visuales reforzarían la confirmación de que sus actos intencionados se están llevando a cabo. Insértese ahora un breve retraso en la cinta de vídeo (uno o dos fotogramas de 33 msegs cada uno), de modo que la confirmación visual siempre se retrase un breve instante de tiempo. Mi predicción es que los sujetos se acostumbrarán rápidamente a ello. (Eso espero, porque el paso siguiente es el interesante). Una vez los sujetos se hayan acostumbrado al retraso, si éste se suprime repentinamente, sentirían como si *el lápiz se doblara,* porque la percepción de la trayectoria de su punta se retrasaría, en relación a la información visual de entrada, como si la punta se arrastrara por la estela de su trayectoria esperada.

4. *Ajustando el retraso en el carrusel de Grey Walter:* es el experimento de continuación que mide el retraso necesario para eliminar el efecto de «carrusel precognitivo». Mi predicción es que será mucho menor de los 300-500 msegs que predice una extensión del modelo estaliniano de Libet.

Sobre modelos de pandemónium de la elección de palabras

¿Cómo podríamos demostrar que «las palabras quieren ser dichas»? ¿Es posible controlar la inspiración? Hasta el momento, los experimentos de Levelt han dado unos resultados negativos sorprendentes (véase nota 2, pág. 255). El tipo de variaciones sobre estos experimentos que me gustaría introducir establecería la posibilidad de un uso «creativo» de las palabras por parte de los sujetos, al tiempo que, de forma muy discreta, se proporcionarían las materias primas necesarias dentro del entorno para que pudieran ser incorporadas por los sujetos en sus producciones. Por ejemplo, se podría preparar a los sujetos para el experimento con dos protocolos diferentes, en los que palabras, un poco sorprendentes, vívidas, ligeramente novedosas o fuera de lugar se liberasen «casualmente» (en pósters colgados en la pared, en las instrucciones para los sujetos, etc.); seguidamente, se daría la oportunidad a los sujetos de expresarse libremente sobre temas en que dichas palabras tuvieran una baja probabilidad de aparecer, de modo que un efecto de *priming* por los preparativos demostraría que dichas expresiones

habrían sido «activadas» y permanecerían al acecho, esperando la oportunidad para ser utilizadas. Si no se detectara ningún efecto, se confirmaría el modelo de Levelt; un efecto muy fuerte (si se produjeran usos muy «forzados») apoyaría un modelo de pandemónium.

Experimentos con rastreadores de ojos

1. *«Visión ciega» en sujetos normales:* los experimentos con rastreadores de ojos en sujetos normales han demostrado que cuando un estímulo en la región parafoveal se altera durante un movimiento sacádico, los sujetos no perciben el cambio (no refieren haber percibido el cambio), pero existen efectos potenciadores: las latencias para la identificación del segundo estímulo se reducen o no, dependiendo de la información recogida a partir del estímulo parafoveal original. Si los sujetos bajo estas condiciones efectúan una conjetura de elección forzada en cuanto a si se alteró el estímulo o no (o si el estímulo inicial era, por ejemplo, una letra mayúscula o una letra minúscula), ¿mostrarán un alto nivel de aciertos? Mi predicción es que sí, para una serie de casos interesantes, pero no mejor que los mejores niveles que se observan en pacientes con visión ciega.

2. *Experimentos con «papel pintado»:* utilizando un rastreador de ojos, y variando los rasgos finos y gruesos de las regiones parafoveales de campos repetitivos de «papel pintado» durante los movimientos sacádicos, organícese la competición necesaria para desestimar la conclusión según la cual hay «más Marilyns». (Puesto que los nuevos resultados de Ramachandran y Gregory me sorprendieron, me la voy a jugar y voy a predecir que *no* hay efectos *graduales* detectables, aunque en los niveles en que los sujetos detectan los cambios es posible que éstos refieran extraños movimientos ilusorios.)

3. *El damero coloreado:* un experimento diseñado para demostrar lo *poco* que hay en el «pleno del campo visual». Se solicita a los sujetos que efectúen una tarea de identificación o interpretación visual que requiere múltiples movimientos sacádicos sobre una escena en movimiento: los sujetos observan unas figuras animadas en blanco y negro proyectadas sobre un fondo de un damero coloreado al azar. Los cuadros son relativamente grandes, por ejemplo, el TRC se divide en una matriz de 12 × 18 cuadros coloreados de modo diferente al azar. (Los colores se escogen al azar, de modo que el patrón no resulta significativo para la tarea visual que se superpone sobre este fondo.) Debería haber diferencias de luminancia entre los cuadros, de modo que no se produzca el efecto Liebmann, y para cada cuadro debería prepararse un *color isoluminante alternativo:* un color que, si sustituyese al color del que originalmente es el cuadro, no creara fronteras de luminancia radicalmente distintas en los bordes (así, los detectores de bordes de luminancia se mantendrán inactivos). Supóngase ahora que, durante los movimientos sacádicos (según se determina gracias al rastreador de ojos), se cambian

los colores del damero; los observadores percibirían que uno o más cuadros cambian de color varias veces por segundo. Predicción: habrá condiciones bajo las cuales los sujetos permanecerán completamente ignorantes al hecho de que ciertas porciones «del fondo» han cambiado bruscamente de color. ¿Por qué? Porque el sistema visual parafoveal es, fundamentalmente, un sistema de alarma, compuesto de centinelas diseñados para provocar un movimiento sacádico cuando se detecta un cambio; un sistema así no se ocuparía de seguir la pista de unos colores insignificantes entre fijaciones y, por tanto, no quedaría nada con lo que comparar el nuevo color. (Ello depende, evidentemente, de lo «rápida que sea la película» en las regiones que responden al color parafoveal; puede que haya un lento período refractario que anule el efecto que predigo.)

BIBLIOGRAFÍA

Akins, K. A. (1989). *On Piranhas, Narcissim and Mental Representation: An Essay on Intentionality and Naturalism,* tesis doctoral inédita, Departamento de Filosofía, University of Michigan, Ann Arbor, MI.

Akins, K. A. (1990). «Science and our inner lives: Birds of prey, bats, and the common (featherless) biped», en Bekoff, M. y D. Jamieson (comps.), *Interpretation and Explanation in the Study of Animal Behavior.* Vol. I, Boulder, CO, Westview, págs. 414-427.

Akins, K. A. y D. C. Dennett (1986). «Who may I say is calling?», *Behavioral and Brain Sciences* 9: 517-518.

Allman, J., F. Meizin y E. L. McGuinness (1985). «Direction and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area», *Perception* 14: 105-126.

Allport, A. (1988). «What concept of consciousness?», en Marcel, A. J. y E. Bisiach (comps.), *Consciousness in Contemporary Science,* Clarendon Press, Oxford, págs. 159-182.

Allport, A. (1989). «Visual attention», en Posner, M. (comp.), *Foundations of Cognitive Psychology,* Cambridge, MA, The MIT Press, págs. 631-682.

Anderson, J. (1983). *The Architecture of Cognition,* Cambridge, MA, Harvard University Press.

Anscombe, G. E. M. (1957). *Intention,* Oxford, Blackwell. [Trad. cast. de Ana Isabel Stellino: *Intención,* Barcelona, Paidós, 1991.]

Anscombe, G. E. M. (1965). «The intentionality of sensation: A grammatical feature», en Butler, R. J. (comp.), *Analytical Philosophy* (segunda serie), Oxford, Blackwell, pág. 160.

Anton, G. (1899). «Ueber die Selbstwahrnehmung der Herderkrankungen des Gehirs durch den Kranben bei Rindenblindheit und Rindentaubheit», *Archiv für Psychiatrie und Nervenkrankheitene* 32: 86-127.

Arnauld, A. (1641). «Quatrièmes résponses aux objections». [Versión inglesa en Cottingham, J., R. Stoofhoff y D. Murdoch, *The Philosophical Writings of Descartes,* Vol. II, 1984, Cambridge, Cambridge University Press.]

Baars, B. (1988). *A Cognitive Theory of Consciousness,* Cambridge, Cambridge University Press.

Bach-y-Rita, P. (1972). *Brain Mechanisms in Sensory Substitution,* Nueva York y Londres, Academic Press.

Ballard, D. y J. Feldman (1982). «Connectionist models and their properties», *Cognitive Science* 6: 205-254.

Bechtel, W. y A. Abrahamsen (1991). *Connectionism and the Mind: An Introduction to Parallel Processing in Networks,* Oxford, Blackwell.

Bennett, J. (1965). «Substance, reality and primary qualities», *American Philosophical Quarterly* 2: 1-17.

Bennett, J. (1976). *Linguistic Behavior,* Cambridge, Cambridge University Press.

Bentham, J. (1789). *Introduction to Principles of Morals and Legislation,* Londres.

Bick, P. A. y M. Kinsbourne (1987). «Auditory hallucinations and subvocal speech in schizophrenic patients», *American Journal of Psychiatry* 144: 222-225.

Bieri, P. (1990). Comentario en el congreso «The Phenomenal Mind - How Is It Possible and Why Is It Necessary?». Zentrum für Interdisziplinäre Forschung, Bielefeld, Alemania, 14-17 de mayo.

Birnbaum, L. y G. Collins (1984). «Opportunistic planning and Freudian slips», *Proceedings, Cognitive Science Society,* Boulder, CO, págs. 124-127.

Bisiach, E. (1988). «The (haunted) brain and consciousness», en Marcel, A. y E. Bisiach (comps.), *Consciousness in Contemporary Science,* Nueva York, Oxford University Press.

Bisiach, E. y C. Luzzatti. (1978). «Unilateral neglect of representational space», *Cortex* 14: 129-133.

Bisiach, E. y G. Vallar (1988). «Hemineglect in humans», en Boller, F. y J. Grafman (comps.), *Handbook of Neuropsychology.* Vol. I, Nueva York, Elsevier.

Bisiach, E., G. Vallar, D. Perani, C. Papagno y A. Berti (1986). «Unawareness of disease following lesions of the right hemisphere: Anosognosia for hemiplegia and anosognosia for hemianopia», *Neuropsychologia* 24: 471-482.

Blakemore, C. (1976). *Mechanics of the Mind,* Cambridge, Cambridge University Press.

Block, N. (1978). «Troubles with functionalism», en Savage, W. (comp.), *Perception and Cognition: Issues in the Foundations of Psychology,* Saint Paul, MN, University of Minnesota Press (Minnesota Studies in the Philosophy of Science IX), págs. 261-326. [Trad. cast. de Eleonora Baringoltz en Rabossi, E.: *Filosofía de la mente y ciencia cognitiva,* Barcelona, Paidós, 1995.]

Block, N. (1981). «Psychologism and Behaviorism», *Philosophical Review* 90: 5-43.

Block, N. (1990). «Inverted Earth», en Tomberlin, J. E. (comp.), *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind, 1990,* Atascadero, CA, Ridgeview Publishing, págs. 53-79.

Boghossian, P. A. y J. D. Velleman (1989). «Colour as a secondary quality», *Mind* 98: 81-103.

Boghossian, P. A. y J. D. Velleman (1991). «Physicalist theories of color», *Philosophical Review* 100: 67-106.

Booth, W. (1988). «Voodoo Science», *Science* 240: 274-277.

Borges, J. L. (1941). «El jardín de senderos que se bifurcan», cap. de *Ficciones,* Madrid, Alianza, [1971], págs. 101-116.

Borgia, G. (1986). «Sexual selection in bowerbirds», *Scientific American* 254: 92-100.

Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology,* Cambridge, MA, The MIT Press (Bradford Books).

Breitmeyer, B. G. (1984). *Visual Masking,* Oxford, Oxford University Press.

Broad, C. D. (1925). *Mind and Its Place in Nature,* Londres, Routledge & Kegan Paul.

Brooks, B. A., J. T. Yates y R. D. Coleman (1980). «Perception of images moving at saccadic velocities during saccades and during fixation», *Experimental Brain Research* 40: 71-78.

Byrne, R. y A. Whiten (comps.) (1988). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans,* Oxford, Clarendon.

Calvanio, R., P. N. Petrone y D. N. Levine (1987). «Left visual spatial neglect is both environment-centered and body-centered», *Neurology* 37: 1.179-1.183.

Calvin, W. (1983). *The Throwing Madonna: Essays on the Brain,* Nueva York, McGraw Hill.

Calvin, W. (1986). *The River that Flows Uphill: A Journey from the Big Bang to the Big Brain,* San Francisco, Sierra Club Books.

Calvin, W. (1987). «The brain as a Darwin Machine», *Nature* 330: 33-34.
Calvin, W. (1989a). *The Cerebral Symphony: Seashore Reflections on the Structure of Consciousness,* Nueva York, Bantam.
Calvin, W. (1989b). «A global brain theory», *Science* 240: 1.802-1.803.
Campion, J., R. Latto y Y. M. Smith (1983). «Is blindsight an effect of scattered light, spared cortex, and near-threshold vision?», *Behavioral and Brain Sciences* 6: 423-486.
Camus, A. (1942). *Le mythe de Sisyphe,* París, Gallimard. [Trad. cast. de Luis Echávarri: *El mito de Sísifo,* Madrid, Alianza, 1951.]
Carruthers, P. (1989). «Brute experience», *Journal of Philosophy* 86: 258-269.
Castañeda, C. (1968). *The Teachings of Don Juan: A Yaqui Way of Knowledge,* Berkeley, CA, University of California Press. [Trad. cast. de Juan Tovar con prólogo de Octavio Paz: *Las enseñanzas de Don Juan. Una forma yaqui de conocimiento,* México, Fondo de Cultura Económica, 1974.]
Castañeda, H.-N. (1967). «Indicators and quasi-indicators», *American Philosophical Quarterly* 4: 85-100.
Castañeda, H.-N. (1968). «On the logic of attributions of self-knowledge to others», *Journal of Philosophy* 65: 439-456.
Changeux, J.-P. y A. Danchin (1976). «Selective stabilization of developing synapses as a mechanism for the specifications of neuronal networks», *Nature* 264: 705-712.
Changeux, J.-P. y S. Dehaene (1989). «Neuronal models of cognitive functions», *Cognition* 33: 63-109.
Cheney, D. L. y R. M. Seyfarth (1990). *How Monkeys See the World,* Chicago, IL, The University of Chicago Press.
Cherniak, C. (1986). *Minimal Rationality,* Cambridge, MA, The MIT Press (Bradford Books).
Churchland, P. M. (1985). «Reduction, qualia and the direct inspection of brain states», *Journal of Philosophy* 82: 8-28.
Churchland, P. M. (1990). «Knowing qualia: A reply to Jackson», cap. 4 de *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science,* Cambridge, MA, The MIT Press (Bradford Books), págs. 67-76.
Churchland, P. S. (1981a). «On the alleged backwards referral of experiences and its relevance to the mind-body problem», *Philosophy of Science* 48: 165-181.
Churchland, P. S. (1981b). «The timing of sensations: Reply to Libet», *Philosophy of Science* 48: 492-497.
Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind/Brain,* Cambridge, MA, The MIT Press (Bradford Books).
Clark, R. W. (1975). *The Life of Bertrand Russell,* Londres, Weidenfeld and Nicolson.
Cohen, L. D., D. Kipnis, E. C. Kunkle y P. E. Kubzansky (1955). «Case report: Observation of a person with congenital insensitivity to pain», *Journal of Abnormal and Social Psychology* 51: 333-338.
Cole, D. (1990). «Functionalism and inverted spectra», *Synthese* 82: 207-222.
Crane, H. y T. P. Piantanida (1983). «On seeing reddish green and yellowish blue», *Science* 222: 1.078-1.080.
Crick, F. (1984). «Function of the thalamic reticular complex: The searchlight hypothesis», *Proceedings of the National Academy of Sciences* 81: 4.5864.590.
Crick, F. y C. Koch (1990). «Towards a neurobiological theory of consciousness», *Seminars in the Neurosciences* 2: 263-275.
Damasio, A. R., H. Damasio y G. W. Van Hoesen (1982). «Prosopagnosia: Anatomic basis and behavioral mechanisms», *Neurology* 32: 331-341.
Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex* 2 Vols., Lon-

dres, Murray. [Trad. cast. parcial de R. Masera en el libro de R. Jastrow. *Charles Darwin. Textos Fundamentales,* Barcelona, Paidós, 1984.]
Davis, W. (1985). *The Serpent and Rainbow,* Nueva York, Simon & Schuster.
Davis, W. (1988a). *Passage of Darkness: The Ethnobiology of the Haitian Zombie,* Chapel Hill, NC y Londres, University of North Carolina Press.
Davis, W. (1988b). «Zombification», *Science* 240: 1.715-1.716.
Dawkins, M. S. (1980). *Animal Suffering: The Science of Animal Welfare,* Londres, Chapman & Hall.
Dawkins, M. S. (1987). «Minding and mattering», en Blakemore, C. y S. Greenfield (comps.), *Mindwaves,* Oxford, Blackwell, págs. 150-160.
Dawkins, M. S. (1990). «From an animal's point of view: Motivation, fitness, and animal welfare», *Behavioral and Brain Sciences* 13: 1-61.
Dawkins, R. (1976). *The Selfish Gene,* Oxford, Oxford University Press, 2ª edición, 1989. [Trad. cast. de J. Robles Suárez y J. Tola Alonso: *El gen egoísta.* Barcelona. Labor, 1979, Salvat, 1988 y 1993.]
Dawkins, R. (1982). *The Extended Phenotype,* San Francisco, CA, Freeman.
Dawkins, R. (1976). *The Blind Watchmaker,* Nueva York, Norton. [Trad. cast.: *El relojero ciego,* Barcelona, Labor, 1989.]
de Sousa, R. (1976). «Rational homunculi», en Rorty, A. O. (comp.), *The Identity of Persons,* Berkeley, CA, University of California Press, págs. 217-238.
Dennett, D. C. (1969). *Content and Consciousness,* Londres, Routledge & Kegan Paul, 3ª edición, 1993.
Dennett, D. C. (1971). «Intentional systems», *Journal of Philosophy* 8: 87-106.
Dennett, D. C. (1974). «Why the law of effect will not go away», *Journal of the Theory of Social Behaviour* 5: 169-187. Reimpreso en Dennett, 1978a.
Dennett, D. C. (1975). «Are dreams experiences?», *Philosophical Review* 73: 151-171.
Dennett, D. C. (1978a). *Brainstorms,* Montgomery, VT, Bradford Books. [Reeditado en 1981 por The MIT Press.]
Dennett, D. C. (1978b). «Skinner skninned», cap. 4 de *Brainstorms,* Montgomery, VT, Bradford Books, págs. 53-70.
Dennett, D. C. (1978c). «Two approaches to mental images», cap. 10 de *Brainstorms,* Montgomery, VT, Bradford Books, págs. 174-189.
Dennett, D. C. (1978d). «Where am I?», cap. 17 de *Brainstorms,* Montgomery, VT, Bradford Books, págs. 310-323.
Dennett, D. C. (1979a). «On the absence of phenomenology», en Gustafson, D. y B. Tapscott (comps.), *Body, Mind and Method: Essays in Honor of Virgil Aldrich,* Dordrecht, Reidel.
Dennett, D. C. (1979b). «Review of Popper and Eccles, *The Self and Its Brain: An Argument for Interactionism*», *Journal of Philosophy* 76: 91-97.
Dennett, D. C. (1981a). «Reflections on "Software" », en Hofstadter, D. R. y D. C. Dennett (comps.), *The Mind's I: Fantasies and Reflections on Self and Soul,* Nueva York, Basic Books, págs. 265-268.
Dennett, D. C. (1981b). «Wondering where the yellow went» (commentary on W. Sellar's Carus Lectures)», *Monist* 64: 102-108.
Dennett, D. C. (1982a). «How to study human consciousness empirically, or nothing comes to mind», *Synthese* 59: 159-180.
Dennett, D. C. (1982b). «Why we think what we do about why we think what we do: Discussion on Goodman's "On thoughts without words"», *Cognition* 12: 219-227.
Dennett, D. C. (1982c). «Comments on Rorty», *Synthese* 59: 349-356.
Dennett, D. C. (1982d). «Notes on prosthetic imagination», *New Boston Review* junio: 3-7.

Dennett, D. C. (1983). «Intentional systems in cognitive ethology: The "Panglossian Paradigm" defended», *Behavioral and Brain Sciences* 6: 343-390.
Dennett, D. C. (1984a). *Elbow Room: The Varieties of Free Will Worth Wanting,* Cambridge, MA, The MIT Press (Bradford Books). [Trad. cast. de Gabriela Ventureira: *La libertad de acción. Un análisis de las exigencias del libre albedrío.* Barcelona, Gedisa, 1992.]
Dennett, D. C. (1984b). «Carving the mind at its joints». Review of J. Fodor, *The Modularity of Mind, Contemporary Psychology* 29: 285-286.
Dennett, D. C. (1985a). «Can machines think?», en Shafto, M. (comp.), *How We Know,* Nueva York, Harper & Row, págs. 121-145.
Dennett, D. C. (1985b). «Music of the hemispheres. A review of M. Gazzaniga, *The Social Brain», New York Times Book Review,* 17 de noviembre, pág. 53.
Dennett, D. C. (1986). «Julian Jaynes' software archeology», *Canadian Psychology* 27: 149-154.
Dennett, D. C. (1987a). *The Intentional Stance,* Cambridge, MA, MIT Press (Bradford Books). [Trad. cast. de Daniel Zadunisky: *La actitud intencional.* Barcelona, Gedisa, 1991.]
Dennett, D. C. (1987b). «The logical geography of computational approaches: A view from the East Pole», en Harnish, M. y M. Brand (comps.), *Problems in the Representation of Knowledge,* Tucson, University of Arizona Press.
Dennett, D. C. (1988a). «Quining qualia», en Marcel, A. y E. Bisiach (comps.), *Consciousness in Contemporary Science,* Nueva York, Oxford University Press, págs. 42-77.
Dennett, D. C. (1988b). «When philosophers encounter AI», *Daedalus* 117: 283-296. Reimpreso en Graubard, 1988.
Dennett, D. C. (1988c). «Out of the armchair and into the field», *Poetics Today* 9: 205-222.
Dennett, D. C. (1988d). «The intentional stance in theory and practice», en Byrne, R. y A. Whiten (comps.), *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans,* Oxford, Clarendon, págs. 180-202.
Dennett, D. C. (1988e). «Science, philosophy and interpretation», *Behavioral and Brain Sciences* 11: 535-546.
Dennett, D. C. (1988f). «Why everyone is a novelist?», *Times Literary Supplement,* 16-22 septiembre, págs. 1.016, 1.028-1.029.
Dennett, D. C. (1989a). «Why creative intelligence is hard to find», *Behavioral and Brain Sciences* 11: 253. Comentario a Whiten y Byrne.
Dennett, D. C. (1989b). «The origins of selves», *Cogito* 2: 163-173.
Dennett, D. C. (1989c). «Murmurs in the Cathedral». Review of R. Penrose, *«The Emperor's New Mind», Times Literary Supplement,* 29 de septiembre-5 de octubre; págs. 1.066-1.068.
Dennett, D. C. (1989d). «Cognitive ethology: Hunting for bargains or a wild goose chase?», en Montefiore, A. y D. Noble (comps.), *Goals, Own Goals and No Goals: A Debate on Goal-Directed and Intentional Behaviour,* Londres, Unwin Hyman.
Dennett, D. C. (1990a). «Memes and the exploitation of imagination», *The Journal of Aesthetics and Art Criticism* 48: 127-135.
Dennett, D. C. (1990b). «Thinking with a computer», en Barlow, H. (comp.), *Image and Understanding,* Cambridge, Cambridge University Press, págs. 297-309.
Dennett, D. C. (1990c). «Betting your life on an algorithm», *Behavioral and Brain Sciences* 13: 660. Comentario a Penrose.
Dennett, D. C. (1990d). «The interpretation of texts, people, and other artifacts», *Philosophy and Phenomenological Research* 50: 177-194.
Dennett, D. C. (1990e). «Two Black boxes: A Fable», manuscrito, Tufts University Center for Cognitive Studies, noviembre.

Dennett, D. C. (1991a). «Real patterns», *Journal of Philosophy* 89: 27-51.
Dennett, D. C. (1991b). «Producing future by telling stories», en Ford, K. M. y Z. Pylyshyn (comps.), *Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence,* Norwood, NJ, Ablex (Ablex Series in Theoretical Issues in Cognitive Science.
Dennett, D. C. (1991c). «Mother Nature versus the Walking Encyclopedia», en Ramsey, W., S. Stich y D. Rumelhart (comps.), *Philosophy and Connectionist Theory,* Hillsdale, NJ, Erlbaum.
Dennett, D. C. (1991d). «Two contrasts: Folk Craft versus Folk Science and Belief versus Opinion», en Greenwood, J. (comp.), *The Future of Folk Psychology: Intentionality and Cognitive Science,* Cambridge, Cambridge University Press.
Dennett, D. C. (1991e). «Granny's campaign for safe science», en Loewer, R. y G. Rey (comps.), *Meaning in Mind. Fodor and His Critics,* Oxford, Blackwell, págs. 87-94.
Dennett, D. C. y M. Kinsbourne (1992). «Time and the observer: The where and when of consciousness in the brain», *Behavioral and Brain Sciences* 15: 183-247.
Descartes, R. (1637). *Discours de la méthode,* París. [Trad. cast. de Manuel García Morente: *Discurso del método - Meditaciones metafísicas,* México, Espasa Calpe Mexicana (Selecciones Austral, 21), 1976.]
Descartes, R. (1641). *Meditationes de prima philosophia,* París, Michel Soly. [Trad. cast. de E. López y M. Graña: *Meditaciones metafísicas y otros textos,* Madrid, Gredos (Biblioteca Hispánica de Filosofía, Clásicos 3), 1987.]
Descartes, R. (1664). *Traité de l'homme,* París. [Trad. cast. de Guillermo Quintás: *El tratado del hombre.* Madrid, Alianza Universidad (Filosofía 663), 1990.]
Descartes, R. (1970). *Philosophical Letters.* Edición a cargo de Kenny, A.: Oxford, Clarendon Press.
Dreyfus, H. (1979). *What Computers Can't Do,* Nueva York, Harper & Row, 2ª edición.
Dreyfus, H. L. y S. E. Dreyfus. (1988). «Making a mind versus modeling the brain: Artificial intelligence back at a branchpoint», en Graubard, S. R. (comp.), *The Artificial Intelligence Debate: False Starts, Real Foundations,* Cambridge, MA, MIT Press.
Eccles, J. C. (1985). «Mental summation: The timing of voluntary intentions by cortical activity», *Behavioral and Brain Sciences* 8: 542-547.
Eco, U. (1990a). «After secret knowledge», *Times Literary Supplement,* 22-28 de junio, págs. 666-678.
Eco, U. (1990b). «Some paranoid readings», *Times Literary Supplement,* 29 de junio - 5 de julio, págs. 694, 706.
Edelman, G. (1987). *Neural Darwinism,* Nueva York, Basic Books.
Edelman, G. (1989). *The Remembered Present: A Biological Theory of Consciousness,* Nueva York, Basic Books.
Efron, R. (1967). «The duration of the present», *Proceedings of the New York Academy of Science* 8: 542-543.
Eldredge, N. y S. J. Gould. (1972). «Punctuated equilibria: An alternative to phyletic gradualism», en Schopf, T. J. M. (comp.), *Models in Paleobiology,* San Francisco, Freeman Cooper, págs. 82-115.
Ericsson, K. A. y H. A. Simon. (1984). *Protocol Analysis: Verbal Reports as Data,* Cambridge, MA, MIT Press (Bradford Books).
Evans, G. (1982). *The Varieties of Reference,* Oxford, Oxford University Press, edición a cargo de John McDowell.
Ewert, J.-P. (1987). «The neuroethology of releasing mechanisms: Prey-catching in toads», *Behavioral and Brain Sciences* 10: 337-405.
Farah, M. J. (1988). «Is visual imagery really visual? Overlooked evidence from neuropsychology», *Psychological Review* 95: 307-317.

Farrell, B. A. (1950). «Experience», *Mind* 59: 170-198.
Fehling, M., B. Baars y C. Fisher (1990). «A functional role of repression in an autonomous, resource-constrained agent», *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society,* Hillsdale, NJ, Erlbaum.
Fehrer, E. y D. Raab (1962). «Reaction time to stimuli masked by metacontrast», *Journal of Experimental Psychology* 63: 143-147.
Feynmann, R. (1985). *Surely You're Joking, Mr. Feynmann!,* Nueva York, Norton. [Trad. cast. de Luis Bou: *¿Está usted de broma, Sr. Feynmann?,* Madrid, Alianza, 1987.]
Finke, R. A., S. Pinker y M. J. Farah (1989). «Reinterpreting visual patterns in mental imagery», *Cognitive Science* 13: 51-78.
Flanagan, O. (1991). *The Science of the Mind,* Cambridge, MA, MIT Press (Bradford Books), 2ª edición.
Florh, H. (1990). «Brain processes and phenomenal consciousness: A new and specific hypothesis». Manuscrito, Zentrum für Interdisziplinäre Forschung, Bielefeld (Alemania). Presentado en el congreso «The Phenomenal Mind — How Is It Possible and Why Is It Necessary?», 14-17 de mayo.
Fodor, J. (1975). *The Language of Thought,* Scranton, PA, Crowell. [Trad. cast. de Jesús Fernández Zulaira y revisión de José E. García-Albea: *El lenguaje del pensamiento,* Madrid, Alianza, 1985.]
Fodor, J. (1983). *The Modularity of Mind,* Cambridge, MA, MIT Press. [Trad. cast. de José Manuel Igoa: *La modularidad de la mente. Un ensayo sobre la psicología...*]
Fodor, J. (1990). *A Theory of Content, and Other Essays,* Cambridge, MA, MIT Press (Bradford Books).
Fodor, J. y Z. Pylyshyn (1988). «Connectionism and cognitive architecture: A critical analysis», *Cognition* 28: 3-71.
Fox, I. (1989). «On the nature and cognitive function of phenomenal content — Part one», *Philosophical Topics* 17: 81-117.
French, R. (1991). «Subcognition and the Turing test», *Mind.*
Freud, S. (1988). «El yo y el ello», en *Los textos fundamentales del psicoanálisis,* Madrid, Alianza, págs. 547-598. Trad. cast. de Luis López Ballesteros, Ramón Rey y Gustavo Dessal.
Freyd, J. (1989). «Dynamic mental representations», *Psychological Review* 94: 427-438.
Fuster, J. M. (1981). «Prefrontal cortex in motor control», en *Handbook of Physiology, Section 1: The Nervous System, Vol. II: Motor Control,* American Physiological Society, págs. 1.149-1.178.
Gardner, H. (1975). *The Shattered Mind,* Nueva York, Knopf.
Gardner, M. (1981). «The Laffer Curve and other laughs in current economics», *Scientific American* 245, diciembre, 18-31. Reimpreso en Gardner, 1986.
Gardner, M. (1986). *Knotted Doughnuts and Other Mathematical Diversions,* San Francisco, CA, Freeman.
Gazzaniga, M. (1978). «Is seeing believing: Notes on clinical recovery», en Finger, S. (comp.), *Recovery from Brain Damage: Research and Theory,* Nueva York, Plenum Press, págs. 409-414.
Gazzaniga, M. (1985). *The Social Brain: Discovering the Newtworks of the Mind,* Nueva York, Basic Books.
Gazzaniga, M. y J. Ledoux (1978). *The Integrated Mind,* Nueva York, Plenum Press.
Geldard, F. A. (1977). «Cutaneous stimulis, vitratory and saltatory», *Journal of Investigative Dermatology* 69: 83-87.
Geldard, F. A. y C. E. Sherrick (1972). «The cutaneous "rabbit": A perceptual illusion», *Science* 178: 178-179.

Geldard, F. A. y C. E. Sherrick (1983). «The cutaneous saltatory area and its presumed neural base», *Perception and Psychophysics* 33: 299-304.
Geldard, F. A. y C. E. Sherrick (1986). «Space, time and touch», *Scientific American* 254: 90-95.
Gert, B. (1965). «Imagination and verifiability», *Philosophical Studies* 16: 44-47.
Geschwind, N. y M. Fusillo (1966). «Color-naming defects in association with alexia», *Archives of Neurology* 15: 137-146.
Gide, A. (1948). *Lex faux monnayeurs*, París, Gallimard.
Goodman, N. (1978). *Ways of Worldmaking*, Hassocks, Sussex, Harvester.
Goody, J. (1977). *The Domestication of the Savage Mind*, Cambridge, Cambridge University Press.
Gould, S. (1980). *The Panda's Thumb*, Nueva York, Norton. [Trad. cast. de Antonio Resines: *El pulgar del panda*. Barcelona, Hermann Blume, 1983.]
Gouras, P. (1984). «Color vision», en Osborn, N. y J. Chader (comps.), *Progress in Retinal Research*. Vol. 3, Londres, Pergamon Press, págs. 227-261.
Graubard, S. R. (1988). *The Artificial Intelligence Debate: False Starts, Real Foundations*, Cambridge, MA, MIT Press. Reimpresión de *Daedalus* 117, invierno de 1988.
Grey Walter, W. (1963). Presentación a la Osler Society. Universidad de Oxford.
Grice, H. P. (1957). «Meaning», *Philosophical Review* 66: 377-388.
Grice, H. P. (1969). «Utterer's meaning and intentions», *Philosophical Review* 78: 147-177.
Hacking, I. (1990). «Signing. Review of Sacks, 1989», *London Review of Books*, 5 de abril, págs. 3-6.
Hampl, P. (1989). «The lax habits of the free imagination», *New York Times Book Review*, 5 de marzo, pág. 1, 37-39. Extraído de P. Hampl, comp., 1989. *The Houghton Mifflin Anthology of Short Fiction*, Boston, MA, Houghton Mifflin.
Hanford, M. (1987). *Where's Waldo?*, Boston, Little, Brown. [Trad. cast.: *¿Dónde está Wally?*, Barcelona, Ediciones B, 1988.]
Hardin, C. L. (1988). *Color for Philosophers: Unweaving the Rainbow*, Indianápolis, IN, Hackett.
Hardin, C. L. (1990). «Color and ilusion. Manuscrito.» Presentado en el congreso «The Phenomenal Mind - How Is It Possible and Why Is It Necessary?» Zentrum für Interdisziplinäre Forschung, Bielefeld, Alemania, 14-17 de mayo.
Harman, G. (1990). «The intrinsic quality of experience», en Tomberlin, J. E. (comp.), *Philosophical Perspectives 4: Action Theory and Philosophy of Mind*, Atascadero, CA, Ridgeview, págs. 31-52.
Harnad, S. (1982). «Consciousness: An afterhought», *Cognition and Brain Theory* 5: 29-47.
Harnad, S. (1989). «Editorial commentary», *Behavioral and Brain Sciences* 12: 183.
Haugeland, J. (comp.) (1981). *Mind Design: Philosophy, Psychology, Artificial Intelligence*, Montgomery, VT, Bradford Books. [Reimpreso por Cambridge, VT, MIT Press, 1981.]
Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*, Cambridge, MA, MIT Press (Bradford Books).
Hawking, S. (1988). *A Brief History of Time*, Nueva York, Bantam. [Trad. cast. de Miguel Ortuño: *Historia del tiempo*, Barcelona, Crítica, 1988.]
Hayes, P. (1979). «The naive physics manifesto», en Michie, D. (comp.), *Expert Systems in the Microelectronic Age*, Edimburgo, Edinburgh University Press.
Hayes-Roth, B. (1985). «A blackboard architecture for control», *Artificial Intelligence* 26: 251-321.
Hebb, D. (1949). *The Organization of Behavior: A Neuropsychological Theory*, Nueva York, Wiley.

Hilbert, D. R. (1987). *Color and Color Perception: A Study in Anthropocentric Realism,* Stanford, CA, Center for the Study of Language and Information.

Hintikka, J. (1962). *Knowledge and Belief,* Ithaca, Cornell University Press. [Trad. cast.: *Saber y creer,* Madrid, Tecnos, 1979.]

Hinton, G. E. y S. J. Nowland. 1987. «How learning can guide evolution», *Complex Systems, I,* Technical Report, CMU-CS-86-128, Carnegie Mellon University, págs. 495-502.

Hobbes, T. (1651). *Leviathan,* París. [Trad. cast. de Carlos Mellizo: *Leviatán.* Madrid, Alianza, 1989.]

Hoffman, R. E. (1986). «What can schizophrenic "voices" tell us?», *Behavioral and Brain Sciences,* págs. 535-548.

Hoffman, R. E. y R. E. Kravitz (1987). «Feedforward action regulation and the experience of will», *Behavioral and Brain Sciences* 10: 782-783.

Hofstadter, D. R. (1981a). «The Turing Test: A coffeehouse conversation», *Scientific American,* mayo. Reimpreso en Hofstadter y Dennett (comps.), 1981, págs. 69-92.

Hofstadter, D. R. (1981b). «Reflections [on Nagel]», en Hofstadter, D. R. y D. C. Dennett (comps.), *The Mind's I: Fantasies and Reflections on Self and Soul,* Nueva York, Basic Books, págs. 403-414.

Hofstadter, D. R. (1983). «The architecture of Jumbo», *Proceedings of the Second Machine Learning Workshop,* Monticello, IL.

Hofstadter, D. R. (1985). «On the seeming paradox of mechanizing creativity», cap. 23 de *Metamagical Themas,* Nueva York, Basic Books, págs. 526-546.

Hofstadter, D. R. y D. C. Dennett (comps.) (1981). *The Mind's I: Fantasies and Reflections of Self and Soul,* Nueva York, Basic Books, págs. 191-201.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems,* Ann Arbor, MI, University of Michigan Press.

Holland, J. H., K. J. Holyoak, R. E. Nisbett y P. R. Thagard (1986). *Induction: Processes of Inference, Learning, and Discovery,* Cambridge, MA, MIT Press (Bradford Books).

Honderich, T. (1984). «The time of a conscious sensory experience and mind-brain theories», *Journal of Theoretical Biology* 110: 115-129.

Howell, R. (1979). «Fictional objects: How they are and how they aren't», en Gustafson, D. F. y B. L. Tapscott (comps.), *Body, Mind and Method,* Dordrecht, D. Reidel, págs. 241-294.

Hughlings Jackson, J. (1915). «Hughlings Jackson on aphasia and kindred affections of speech», *Brain* 38: 1-190.

Hume, D. (1739). *Treatise on Human Nature,* Londres, John Noon. [Trad. cast. de Félix Duque: *Tratado de la naturaleza humana.* Madrid, Editora Nacional, 1980-1981 (2 vols.) y (traducción revisada) Madrid, Tecnos, 1988.]

Humphrey, N. (1972). «"Interest" and "pleasure": Two determinants of a monkey's visual preferences», *Perception* 1: 395-416.

Humphrey, N. (1976). «The colour currency of nature», en Porter, T. y B. Mikellides (comps.), *Colour for Architecture,* Londres, Studio-Vista, págs. 147-161. Reimpreso en Humphrey, 1983a.

Humphrey, N. (1983a). *Consciousness Regained,* Oxford, Oxford University Press.

Humphrey, N. (1983b). «The adaptiveness of mentalism?» Comentario sobre Dennett 1983, *Behavioral and Brain Sciences* 6: 366.

Humphrey, N. (1986). *The Inner Eye,* Londres, Faber & Faber.

Humphrey, N. (1992). *A History of the Mind,* Nueva York, Chatto & Windus.

Humphrey, N. y D. C. Dennett (1989). «Speaking for our selves: An asessment of Multiple Personality Disorder», *Raritan* 9: 68-98.

Humphrey, N. y G. Keeble (1978). «Effects of red light and loud noise on the rates at which monkeys sample the sensory environment», *Perception* 7: 343.
Hundert, E. (1987). «Can neuroscience contribute to philosophy?», en Blakemore, C. y S. Greenfield (comps.), *Mindwaves*, Oxford, Blackwell, págs. 407-429. Reimpreso como el capítulo 7 de E. Hundert, *Philosophy, Psychiatry, and Neuroscience: Three Approaches to the Mind*, Oxford, Clarendon, 1989.
Huxley, T. (1874). «On the hypotesis that animals are automata», en *Collected Essays*, Londres (1893-1894).
Jackendoff, R. (1987). *Consciousness and the Computational Mind*, Cambridge, CA, MIT Press (Bradford Books).
Jackson, F. (1982). «Epiphenomenal qualia», *Philosophical Quarterly* 32: 127-136.
Jacob, F. (1982). *The Possible and the Actual*, Seattle, WA, University of Washington Press.
Janlert, L.-E. (1985). *Studies in Knowledge Representation*, Umea, Suecia, Institute of Information Processing.
Jarrell, R. (1963). *The Bat-Poet*, Nueva York, Macmillan.
Jaynes, J. (1976). *The origins of Consciousness in the Breakdown of the Bicameral Mind*, Boston, MA, Houghton Mifflin.
Jerison, H. (1973). *Evolution of the Brain and Intelligence*, Nueva York, Academic Press.
Johnson-Laird, P. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*, Cambridge, Cambridge University Press.
Johnson-Laird, P. (1988). «A computational analysis of consciousness», en Marcel, A. J. y E. Bisiach (comps.), *Consciousness in Contemporary Science*, Oxford, Clarendon Press.
Julesz, B. (1971). *Foundations of Cyclopean Perception*, Chicago, The University of Chicago Press.
Keller, H. (1908). *The World I Live In*, Nueva York, Century Co.
Kinsbourne, M. (1974). «Lateral interactions in the brain», en Kinsbourne, M. y W. L. Smith (comps.), *Hemisphere Disconnection and Cerebral Function*, Springfield, IL, Charles C. Thomas, págs. 239-259.
Kinsbourne, M. (1980). «Brain-based limitations on mind», en Rieber, R. W. (comps.), *Body and Mind: Past, Present and Future*, Nueva York, Academic Press, págs. 155-175.
Kinsbourne, M. y R. E. Hicks. (1978). «Functional cerebral space: A model for overflow, transfer and interference effects in human performance: A tutorial review», en Requin, J. (comp.), *Attention and Performance 7*, Hillsdale, NJ, Erlbaum, págs. 345-362.
Kinsbourne, M. y W. L. Smith (comps.) (1974). *Hemisphere Disconnection and Cerebral Function*, Springfield, IL, Charles C. Thomas.
Kinsbourne, M. y E. K. Warrington. (1963). «Jargon aphasia», *Neuropsychologia* 1: 27-37.
Kirman, B. H. y otros (1968). «Congenital insensitivity to pain in an imbecile boy», *Developmental Medicine and Child Neurology* 10: 57-63.
Kitcher, P. (1979). «Phenomenal qualities», *American Philosophical Quarterly* 16: 123-129.
Koestler, A. (1967). *The Ghost in the Machine*, Nueva York, Macmillan.
Kohler, I. (1961). «Experiments with goggles», *Scientific American* 206: 62-86.
Kolers, P. A. (1972). *Aspects of Motion Perception*, Londres, Pergamon Press.
Kolers, P. A. y M. von Grünau (1976). «Shape and color in apparent motion», *Vision Research* 16: 329-335.

Kosslyn, S. M. (1980). *Image and Mind*, Cambridge, MA, Harvard University Press.
Kosslyn, S. M., J. D. Holtzman, M. S. Gazzaniga y M. J. Farah (1985). «A computational analysis of mental imagery generation: Evidence for functional dissociation in split brain patients», *Journal of Experimental Psychology: General*, 114: 311-341.
Lackner, J. R. (1988). «Some proprioceptive influences on the perceptual representation of body shape and orientation», *Brain* 111: 281-297.
Langton, C. G. (1989). *Artificial Life*, Redwood City, CA, Addison-Wesley.
Larkin, S. y H. A. Simon. (1987). «Why a diagram is (sometimes) worth ten thousand words», *Cognitive Science* 11: 65-100.
Leiber, J. (1988). «Cartesian' linguistics?», *Philosophia* 118: 309-346.
Leiber, J. (1991). *Invitation to Cognitive Science*, Oxford, Blackwell.
Leibniz, G. W. (1840). «Monadology», en Erdmann, J. E. (comp.), *Leibniz, Opera Philosophica 2*, Berlín. [Trad. cast. de la E. de Olaso: «Monadología», en *Escritos Filosóficos*, Buenos Aires, Ed. Charcas, 1982.]
Levelt, W. (1989). *Speaking*, Cambridge, MA, MIT Press (Bradford Books).
Levy, J. y C. Trevarthen (1976). «Metacontrol of hemispheric function in human split-brain patients», *Journal of Experimental Psychology: Human Perception and Performance* 3: 299-311.
Lewis, D. (1978). «Truth in fiction», *American Philosophical Quarterly* 15: 37-46.
Lewis, D. (1979). «Attitudes *De Dicto* and *De Se*», *Philosophical Review* 78: 513-543.
Lewis, D. (1988). «What experience teaches». Proceedings of the Russellian Society of the University of Sydney. [Reimpreso en W. Lycan, comp., *Mind and Cognition: a Reader*, Oxford, Blackwell, 1990.]
Liberman, A. y M. Studdert-Kennedy (1977). «Phonetic perception», en Held, R., H. Leibowitz y H.-L. Teuber (comps.), *Handbook of Sensory Physiology*. Vol. 8, *Perception*, Heidelberg, Springer Verlag.
Libet, B. (1965). «Cortical activation in conscious and unconscious experience», *Perspectives in Biology and medicine* 9: 77-86.
Libet, B. (1981). «The experimental evidence for subjective referral of a sensory experience backwards in time: Reply to P. S. Churchland», *Philosophy of Science* 48: 182-197.
Libet, B. (1982). «Brain stimulation in the study of neuronal functions for conscious sensory experiences», *Human Neurobiology* 1: 235-242.
Libet, B. (1985a). «Unconscious cerebral initiative and the role of conscious will in voluntary action», *Behavioral and Brain Sciences* 8: 529-566.
Libet, B. (1985b). «Subjective antedating of a sensory experience and mind-brain theories», *Journal of Theoretical Biology* 114: 563-570.
Libet, B. (1987). «Are mental experiences of will and self-control significant for the performance of a voluntary act?», *Behavioral and Brain Sciences* 10: 783-786.
Libet, B. (1989). «The timing of a subjective experience», *Behavioral and Brain Sciences* 12: 183-185.
Libet, B., E. W. Wright, B. Feinstein y D. K. Pearl (1979). «Subjective referral of the timing for a conscious sensory experience», *Brain* 102: 193-224.
Liebmann, S. (1927). «Ueber das Verhalten fahrbiger Formen bei Heligkeitsgleichtheit von Figur und Grund», *Psychologie Forschung* 9: 200-253.
Livingstone, M. S. y D. H. Hubel (1987). «Psychophysical evidence for separate channels for the perception of form, color, movement, and depth», *Journal of Neuroscience* 7: 346-368.
Lloyd, M. y H. S. Dybas (1966). «The periodical cicada problem», *Evolution* 20: 132-149.
Loar, B. (1990). «Phenomenal properties», en Tomberlin, J. E. (comp.), *Philosophical*

Perspectives 4: Action Theory and Philosophy of Mind, 1990, Atascadero, CA, Ridgeview Publishing, págs. 81-108.
Locke, J. (1690). *Essay Concerning Human Understanding,* Londres, Basset. [Trad. cast. de Edmundo O'Gorman: *Ensayo sobre el entendimiento humano,* México, Fondo de Cultura Económica, 1956.]
Lockwood, M. (1989). *Mind, Brain and the Quantum,* Oxford, Blackwell.
Lodge, D. (1988). *Nice Work,* Londres, Secker and Warburg.
Lycan, W. (1973). «Inverted spectrum», *Ratio* 15: 315-319.
Lycan, W. (1990). «What is the subjectivity of the mental?», en Tomberlin, J. E. (comp.), *Philosophical Perspectives 4: Action Theory and Philosophy of Mind,* 1990, Atascadero, CA, Ridgeview Publishing, págs. 109-130.
Marais, E. N. (1937). *The Soul of the White Ant,* Londres, Methuen.
Marcel, A. (1988). «Phenomenal experience and funcionalism», en Marcel, A. y E. Bisiach (comps.), *Consciousness in Contemporary Science,* Nueva York, Oxford University Press, págs. 121-158.
Marcel, A. (en prensa) «Slippage in the unity of consciousness», en Bornstein, R. y T. Pittman (comps.), *Perception Without Awareness: Cognitive, Clinical and Social Perspectives,* Nueva York, Guilford Press.
Marcel, A. y E. Bisiach (comps) (1988). *Consciousness in Contemporary Science,* Nueva York, Oxford University Press.
Margolis, H. (1987). *Patterns, Thinking, and Cognition,* Chicago, IL, The University of Chicago Press.
Margulis, L. (1970). *The Origin of Eukaryotic Cells,* New Haven, Yale University Press.
Marks, C. (1980). *Commissurotomy, Consciousness and Unity of Mind,* Cambridge, MA, MIT Press (Bradford Books).
Marler, P. y V. Sherman (1983). «Song structure without auditory feedback: Emendations of the auditory template hypothesis», *Journal of Neuroscience* 3: 517-531.
Marr, D. (1982). *Vision,* San Francisco, Freeman.
Maynard Smith, J. (1978). *The Evolution of Sex,* Cambridge, Cambridge University Press.
Maynard Smith, J. (1989). *Sex, Games, and Evolution,* Brighton, Sussex, Harvester.
McClelland, J. y D. Rumelhart (comps.) (1986). *Parallel Distributed Processing: Explorations in the Microstructures of Cognition,* vol. 2: *Psychological and Biological Models,* Cambridge, MA, MIT Press.
McCulloch, W. S. y W. Pitts (1943). «A logical calculus for the ideas immanent in nervous activity», *Bulletin of Mathematical Biophysics* 5: 115-113.
McGinn, C. (1989). «Can we solve the mind-body problem?», *Mind* 98: 349-366.
McGinn, C. (1990). *The Problem of Consciousness,* Oxford, Blackwell.
McGlynn, S. M. y D. L. Schacter (1989). «Unawareness of deficits in neuropsychological syndromes», *Journal of Clinical and Experimental Neuropsychology* 11: 143-205.
McGurk, H. y R. Macdonald (1979). «Hearing lips and seeing voices», *Nature* 264: 746-748
McLuhan, M. (1967). *The Medium Is the Massage,* Nueva York, Bantam. [Trad. cast. de León Mirlas: *El medio es el masaje,* Barcelona, Paidós, 1988.]
Mellor, H. (1981). *Real Time,* Cambridge, Cambridge University Press.
Menzel, E. W., E. S. Savage-Rumbaugh y J. Lawson (1985). «Chimpanzee (*Pan troglodytes*) spatial problem solving with the use of mirrors and televised equivalents of mirrors», *Journal of Comparative Psychology* 99: 211-217.
Millikan, R. (1990). «Truth rules, hoverflies, and the Kripke-Wittgenstein paradox», *Philosophical Review* 99: 323-354.
Minsky, M. (1975). *A Framework for Representing Knowledge,* Memo 3306, AI Lab MIT. Algunos fragmentos han sido publicados en Haugeland, 1981, pág. 95-128.

Minsky, M. (1985). *The Society of Mind,* Nueva York, Simon & Schuster. [Trad. Cast. de Lidia Espinosa de Matheu: *La sociedad de la mente.* Buenos Aires, Ediciones Galápago, 1986.]

Mishkin, M., L. G. Ungerledier y K. A. Macko (1983). «Object vision and spatial vision: Two cortical pathways», *Trend in Neuroscience* 64: 370-375.

Monod, J. (1969). *Le hasard et la nécessité,* París, Seuil. [Trad. cast. de F. Ferrer Lerín: *El azar y la necesidad.* Barcelona, Barral, 1970 y Barcelona, Tusquets, 1984.]

Morris, R. K., K. Rayner y A. Pollatsek (1990). «Eye movement guidance in reading: The role of parafoveal and space information», *Journal of Experimental Psychology: Human Perception and Performance* 16: 268-281.

Mountcastle, V. B. (1978). «An organizing principle for cerebral function: The unit module and the distributed system», en Edelman, G. y V. B. Mountcastle (comps.), *The Mindful Brain,* Cambridge, MA, MIT Press, págs. 7-50.

Nabokov, V. (1930). «Zaschita Luzhina», en *Sovremennye Zapiski,* París. Publicado en forma de libro por Slovo, Berlín, 1930.

Nagel, T. (1971). «Brain bisection and the unity of consciousness», *Synthese* 22: 396-413. Reimpreso en su libro *Mortal Questions,* Cambridge, Cambridge University Press, 1979.

Nagel, T. (1974). «What Is It Like to Be a Bat?», *Philosophical Review* 83: 435-450.

Nagel, T. (1986). *The View from Nowhere,* Oxford, Oxford University Press.

Neisser, U. (1967). *Cognitive Psychology,* Nueva York, Appleton-Century-Crofts.

Neisser, U. (1981). «John Dean's memory: A case study», *Cognition* 9: 1-22.

Neisser, U. (1988). «Five kinds of self-knowledge», *Philosophical Psychology* 1: 35-39.

Nemirow, L. (1990). «Psysicalism and the cognitive role of acquaintance», en Lycan, W. (comp.), *Mind and Cognition: A Reader,* Oxford, Blackwell, págs. 490-499.

Neumann, O. (1990). «Some aspects of phenomenal consciousness and their possible functional correlates.» Manuscrito. Presentado en el congreso «The Phenomenal Mind - How Is It Possible and Why Is It Necessary?», Zentrum für Interdisziplinäre Forschung, Bielefeld, Alemania, 14-17 de mayo.

Newell, A. (1973). «Production systems: Models of control structures», en Chase, W. G. (comp.), *Visual Information Processing,* Nueva York, Academic Press, págs. 463-526.

Newell, A. (1982). «The knowledge level», *Artificial Intelligence* 18: 81-132.

Newell, A. (1988). «The intentional stance and the knowledge level», *Behavioral and Brain Sciences* 11: 520-522.

Newell, A. (1990). *Unified Theories of Cognition,* Cambridge, MA, Harvard University Press.

Newell, A., P. S. Rosenbloom y J. E. Laird (1989). «Symbolic architectures for cognition», en Posner, M. (comps.), *Foundation for Cognitive Science,* Cambridge, MA, MIT Press, págs. 93-132.

Nielsen, T. I. (1963). «Volition: A new experimental approach», *Scandinavian Journal of Psychology* 4: 225-230.

Nilsson, N. (1984). *Shakey the Computer,* SRI Technical Report, SRI International, Menlo Park, C.A.

Norman, D. A. y T. Shallice (1980). *Attention to Action: Willed and Automatic Control of Behavior,* Technical Report n. 99, Center for Human Information Processing. Reimpreso con revisiones en R. J. Davidson, G. E. Schwartz y D. Shapiro (comps.), 1986. *Consciousness and Self-Regulation.* Nueva York, Plenum Press.

Norman D. A. y T. Shallice (1985). «Attention to action», en Shallice, T. (comp.), *Consciousness and Self-Regulation,* Nueva York, Plenum Press.

Nottebohm, F. (1984). «Birdsong as a model in which to study brain processes related to learning», *Condor* 86: 227-236.
Oakley, D. A. (comp.) (1985). *Brain and Mind,* Londres y Nueva York, Methuen.
Ornstein, R. y R. F. Thompson (1984). *The Amazing Brain,* Boston, MA, Houghton Mifflin.
Pagels H. (1988). *The Dreams of Reason: The Computer and the Rise of the Sciences of Complexity,* Nueva York, Simon & Schuster.
Papert, S. (1988). «One AI or many?», *Daedalus* 117: 1-14.
Parfit, D. (1984). *Reasons and Persons,* Oxford, Clarendon Press.
Pears, D. (1984). *Motivated Irrationality,* Oxford, Clarendon Press.
Penfield, W. (1958). *The Excitable Cortex in Conscious Man,* Liverpool, Liverpool University Press.
Penrose, R. (1989). *The Emperor's New Mind,* Oxford, Oxford University Press. [Trad. cast. de Javier García Sanz: *La nueva mente del emperador,* Madrid, Mondadori (Biblioteca Mondadori, 20), 1991.]
Perlis, D. (1991). «Intentionality and defaults», en Ford, K. M. y P. J. Hayes (comps.), *Reasoning Agents in a Dynamic World,* Greenwich, CT, JAI Press.
Perry, J. (1979). «The Problem of the essential indexical», *Nous* 13: 3-21.
Pinker, S. y P. Bloom (1990). «Natural language and natural selection», *Behavioral and Brain Sciences* 13: 707-784.
Pollatsek, A., K. Rayner y W. E. Collins. (1984). «Integrating pictorial information across eye movements», *Journal of Experimental Psychology: General* 113: 426-442.
Pöppel, E. (1985). *Grenzen des Bewusstseins,* Stuttgart, Deutsche Verlags-Anstal.
Pöppel, E. (1988). *Mindworks: Time and Conscious Experience,* Nueva York, Harcourt Brace Jovanovich. [Traducción de Pöppel, 1985.]
Popper, K. R. y J. C. Eccles (1977). *The Self and Its Brain,* Berlín, Springer Verlag. [Trad. cast. de C. Solís Santos: *El yo y su cerebro,* Barcelona, Labor, 1980.]
Powers, L. (1978). «Knowledge by deduction», *Philosophical Review* 87: 337-371.
Putnam, H. (1965). «Brains and behavior», en Butler, R. J. (comps.), *Analytical Philosophy. Second Series,* Oxford, Blackwell, págs. 1-19.
Putnam, H. (1988). «Much ado about not very much», *Daedalus* 117. Reimpreso en Graubard, 1988.
Pylyshyn, Z. (1979). «Do Mental events have durations?», *Behavioral and Brain Sciences* 2: 277-278.
Quine, W. V. O. (1969). «Natural kinds», en *Ontological Relativity and Other Essays,* Nueva York, Columbia University Press, págs. 114-138.
Ramachandran, V. S. (1985). Editorial en *Perception* 14: 97-103.
Ramachandran, V. S. (1991). «2-D or not 2-D: That is the question», en Gregory, R. L., J. Harris, P. Heard, D. Rose y C. Cronly-Dillon (comps.), *The Artful Brain,* Oxford, Oxford University Press.
Ramachandran, V. S. y R. L. Gregory. s.f. «Perceptual filling in of artificially induced scotomas in human vision.» Manuscrito. Presentado para publicación en *Nature.*
Ramsey, W., S. Stich y D. Rumelhart (comps.) (1991). *Philosophy and Connectionist Theory,* Hillsdale, NJ, Erlbaum.
Raphael, B. (1976). *The Thinking Computer: Mind Inside Matter,* San Francisco, Freeman.
Reddy, D. R., L. D. Erman, R. D. Fennel y R. B. Neely (1973). «The HEARSAY-II speech understanding system: An example of the recognition process», *Proceedings of the International Joint Conference on Artificial Intelligence,* Stanford CA: 185-194.

Reingold, E. M. y P. M. Merikle. 1990. «On the interrelatedness of theory and measurement in the study of unconscious processes», *Mind and Language* 5: 9-28.
Reisberg, D. y D. Chamberss, en preparación. «Neither pictures nor propositions: What can we learn from a mental image?», *Canadian Journal of Psychology.*
Richards, R. J. (1987). *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior,* Chicago, IL, The University of Chicago Press.
Ristau, C. (1991). *Cognitive Ethology: The Minds of Other Animals: Essays in Honor of Donald R. Griffin,* Hillsdale, NJ, Erlbaum.
Rizzolati, G., M. Gentilucci y M. Matelli (1985). «Selective spatial attention: Once center, one circuit, or many circuits?», en Posner, M. I. y O. S. M. Marin (comps.), *Attention and Performance XI,* Hillsdale, NJ, Erlbaum.
Rorty, R. (1970). «Incorrigibility as the mark of the mental», *Journal of Philosophy* 67: 399-424.
Rorty, R. (1982a). «Contemporary philosophy of mind», *Synthese* 53: 323-348.
Rorty, R. (1982b). «Comments on Dennett», *Synthese* 53: 181-187.
Rosenbloom, P. S., J. E. Laird y A. Newell (1987). «Knowledge-level learning in Soar», *Proceedings of AAAI,* Los Altos, CA, Morgan Kaufman.
Rosenthal, D. (1986). «Two concepts of consciousness», *Philosophical Studies* 49: 329-359.
Rosenthal, D. (1989). *Thinking that one thinks,* ZIF, Report n. 11, Research Group on Mind and Brain, Perspectives in Theoretical Psychology and the Philosophy of Mind, Zentrum für Interdisziplinäre Forschung, Bielefeld, Alemania.
Rosenthal, D. (1990a). *Why are verbally expressed thoughts conscious?,* ZIF, Report n. 32, Research Group on Mind and Brain, Perspectives in Theoretical Psychology and the Philosophy of Mind, Zentrum für Interdisziplinäre Forschung, Bielefeld, Alemania.
Rosenthal, D. (1990b). *A theory of consciousness,* ZIF, Report n. 40, Research Group on Mind and Brain, Perspectives in Theoretical Psychology and the Philosophy of Mind, Zentrum für Interdisziplinäre Forschung, Bielefeld, Alemania.
Rozin, P. (1976). «The evolution of intelligence and access to the cognitive unconscious», *Progress in Psychobiology and Physiological Psychology* 6: 245-280.
Rozin, P. (1982). «Human food selection: The interaction of biology, culture and individual experience», en Barker, L. M. (comp.), *The Psychobiology of Human Food Selection,* Westport, CT, Avi Publishing Co.
Rozin, P. y A. Fallon (1987). «A perspective on disgust», *Psychological Review* 94: 23-47.
Rumelhart, D. y J. McClelland (comps.) (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition 1: Foundations,* Cambridge, MA, MIT Press.
Russell, B. (1927). *The Analysis of Matter,* Londres, Allen and Unwin. [Trad. cast.: *Análisis de la materia,* Madrid, Taurus, [2]1976.]
Ryle, G. (1949). *The Concept of Mind,* Londres, Hutchinson.
Ryle, G. (1979). *On Thinking.* Edición a cargo de Kolenda, K.: Totowa, NJ, Rowman and Littlefield.
Sacks, O. (1985). *The Man Who Mistook His Wife for His Hat,* Nueva York, Summit Books. [Trad. cast. de José Manuel Álvarez Flórez: *El hombre que confundió a su mujer con un sombrero,* Barcelona, Muchnik, 1987.]
Sacks, O. (1989). *Seeing Voices,* Berkeley, CA, University of California Press. [Trad. cast. de José Manuel Álvarez Flórez: *«Veo una voz» Viaje al mundo de los sordos,* Madrid, Anaya & Mario Muchnik, 1991.]
Sandeval, E. (1991). «Towards a logic of dynamic frames», en Ford, K. M. y J. Hayes (comps.), *Reasoning Agents in a Dynamic World,* Greenwich, CT, JAI Press.

Sanford, D. (1975). «Infinity and vagueness», *Philosophical Review* 84: 520-535.
Sartre, J.-P. (1943). *L'être et le néant*, París, Gallimard. [Trad. cast. de Juan Valmar: *El ser y la nada*. Buenos Aires, Losada, 1966 - Madrid, Alianza, 1984.]
Schank, R. (1991). *Tell Me a Story*, Nueva York, Scribners.
Schank, R. y R. Abelson (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*, Hillsdale, NJ, Erlbaum. [Trad. cast. de Elisabeth Gilboy y Javier Zanón: *Guiones, planes, metas y entendimiento*, Barcelona, Paidós, 1987.]
Schull, J. (1990). «Are species intelligent?», *Behavioral and Brain Sciences* 13: 63-108.
Searle, J. (1980). «Minds, brains, and programs», *Behavioral and Brain Sciences* 3: 417-458.
Searle, J. (1982). «The myth of the computer: An exchange», *New York Review of Books*, 24 de junio, págs. 56-57.
Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*, Cambridge, Cambridge University Press.
Searle, J. (1984). «Panel discussion: Has artificial intelligence research illuminated human thinking?», en Pagels, H. (comp.), *Computer Culture: The Scientific, Intellectual, and Social Impact of the Computer*, (Annals of the New York Academy of Sciences 426), Nueva York.
Searle, J. (1988a). «Turing the Chinese Room», en Singh T. (comp.), *Synthesis of Science and Religion, Critical Essays and Dialogues*, San Francisco, CA, Bhaktivedenta Institute.
Searle, J. (1988b). «The realistic stance», *Behavioral and Brain Sciences* 11: 527-529.
Searle, J. (1990a). «Consciousness, explanatory inversion, and cognitive science», *Behavioral and Brain Sciences* 13: 585-642.
Searle, J. (1990b). «Is the brain's mind a computer program?», *Scientific American* 262: 26-31. [Trad. cast.: «¿Es la mente un programa informático?», *Investigación y Ciencia*, marzo, 1990, pág. 10-16.]
Selfridge, O. 1959. «Pandemonium: A paradigm for learning», *Symposium on the Mechanization of Thought Processes*, Londres, HM Stationery Office.
Selfridge, O. no publicado. *Tracking and Trailing*.
Sellars, W. (1963). «Empiricism and the Philosophy of Mind», en *Science, Perception and Reality*, Londres, Routledge & Kegan Paul.
Sellars, W. (1981). «Foundations for a metaphysics of pure process», (The Carus Lectures), *Monist* 64: 3-90.
Shallice, T. (1972). «Dual functions of consciousness», *Psychological Review* 79: 383-393.
Shallice, T. (1978). «The dominant action system: An information-processing approach to consciousness», en Pope, K. S. y J. L. Singer (comps.), *The Stream of Consciousness*, Nueva York, Plenum, págs. 148-164.
Shallice, T. (1988). *From Neurophsychology to Mental Structure*, Cambridge, Cambridge University Press.
Sharpe, T. (1977). *The Great Pursuit*, Londres, Secker and Warburg. [Trad. cast.: *La gran pesquisa*, Barcelona, Anagrama, [3]1992.]
Shepard, R. N. (1964). «Circularity in judgments of relative pitch», *Journal of the Acoustical Society of America* 36: 2.346-2.353.
Shepard, R. N. y L. A. Cooper (1982). *Mental Images and Their Transformations*, Cambridge, MA, MIT Press (Bradford Books).
Shepard, R. N. y J. Metzler (1971). «Mental rotation of three-dimensional objects», *Sciences* 171: 701-703.
Shoemaker, S. (1969). «Time without change», *Journal of Philosophy* 66: 363-381.
Shoemaker, S. (1975). «Functionalism and qualia», *Synthese* 27: 291-315.

Shoemaker, S. (1981). «Absent qualia are impossible - A reply to Block», *Philosophical Review* 90: 581-599.
Shoemaker, S. (1988). «Qualia and consciousness». Coloquio en el Departamento de Filosofía de la Tufts University.
Siegel, R. K. y L. J. West (comps.) (1975). *Hallucinations: Behavior, Experience and Theory,* Nueva York, Wiley.
Simon, H. A. y C. A. Kaplan (1989). «Foundations of cognitive science», en Posner, M. I. (comps.), *Foundations of Cognitive Science,* Cambridge, MA, MIT Press.
Smolensky, P. (1988). «On the proper treatment of connectionism», *Behavioral and Brain Sciences* 11: 1-74.
Smullyan, R. M. (1981). «An epistemological nightmare», en Hofstadter, D. R. y D. C. Dennett (comps.), *The Mind's I: Fantasies and Reflections on Self and Soul,* Nueva York, Basic Books, págs. 415-427. Reimpreso en Smullgan, R. M. 1982. *Philosophical Fantasies,* Nueva York, St. Martin's Press.
Smythies, J. R. (1954). «Analysis of projection», *British Journal of Philosophy of Science* 5: 120-133.
Snyder, D. M. (1988). «On the time of a conscious peripheral sensation», *Journal of Theoretical Biology* 130: 253-254.
Sperber, D. y D. Wilson. (1986). *Relevance: A Theory of Communication,* Cambridge, MA, Harvard University Press.
Sperling, G. (1960). «The information available in brief visual presentations», *Psychological Monographs* 74, n. 11.
Sperry, R. W. (1977). «Forebrain commissurotomy and conscious awareness», *The Journal of Medicine and Philosophy* 2: 101-126.
Spillman, L. y J. S. Werner (1990). *Visual Perception: The Neurophysiological Foundations,* San Diego, CA, Academic Press.
Spinoza, B. (1677). *Tractatus de intellectus emendatione.* Edición y traducción de Atilano Domínguez: *Tratado de la reforma del entendimiento,* Madrid, Alianza, 1988.
Stafford, S. P. (1983). *On the Origin of the Intentional Stance,* Working Paper in Cognitive Science CCM 83-1, Tufts University.
Stalnaker, R. (1984). *Inquiry,* Cambridge, MA, MIT Press (Bradford Books).
Stix, G. (1991). «Reach out», *Scientific American* 264: 134.
Stoerig, P. y A. Cowey (1990). «Wavelenght sensitivity in blindsight», *Nature* 342: 916-918.
Stoll, C. (1989). *The Cuckoo's Egg: Tracking a Spy Through the Maze of Computer Espionage,* Nueva York, Doubleday.
Straight, H. S. (1976). «Comprehension versus production in linguistic theory», *Foundations of Language* 14: 525-540.
Stratton, G. M. (1896). «Some preliminary experiments on vision without inversion of the retinal image», *Psychology Review* 3: 611-617.
Strawson, G. (1989). «Red and "Red"», *Synthese* 78: 193-232.
Strawson, P. F. (1962). «Freedom and resentment», *Proceedings of the British Academy.* Reimpreso en P. F. Strawson (comp.), *Studies in the Philosophy of Thought and Action,* Oxford, Oxford University Press, 1968. [Trad. cast. de Juan José Acero: *Libertad y resentimiento,* Barcelona, Paidós, 1995.]
Taylor, D. M. (1966). «The incommunicability of content», *Mind* 75: 527-541.
Thompson, D. W. (1917). *On Growth and Form,* Cambridge, Cambridge University Press. [Trad. cast.: *Sobre el crecimiento y la forma.* Madrid, Blume, 1980.]
Thompson, E., A. Palacios y F. Varela (1992). «Ways of coloring: Comparative color vision as a case study for cognitive science», *Behavioral and Brain Sciences* 15: 1-74.
Tranel, D. y A. R. Damasio (1988). «Non-conscious face recognition in patients with face agnosia», *Behavioral and Brain Research* 30: 235-249.

Tranel, D., A. R. Damasio y H. Damasio (1988). «Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity», *Neurology* 38: 690-696.
Treisman, A. (1988). «Features and objects: The Fourteenth Bartlett Memorial Lecture», *Quarterly Journal of Experimental Psychology* 40A: 201-237.
Treisman, A. y G. Gelade (1980). «A feature-integration theory of attention», *Cognitive Psychology* 12: 97-136.
Treisman, A. y S. Sato (1990). «Conjunction search revisited», *Journal of Experimental Psychology: Human Perception and Performance* 16: 459-478.
Treisman, A. y J. Souther (1985). «Search asymmetry: A diagnostic for preattentive processing of separable features», *Journal of Experimental Psychology: General* 114: 285-310.
Turing, A. (1950). «Computing machinery and intelligence», *Mind* (95): 433-460. [Trad. cast.: «Maquinaria de cómputo e inteligencia». En Z. W. Pylyshyn (comp.), *Perspectivas de la revolución de los computadores,* Madrid, Alianza, 1975.]
Tye, M. (1986). «The sujective qualities of experience», *Mind* 95: 1-17.
Uttal, W. R. (1979). «Do central nonlinearities exist?», *Behavioral and Brain Sciences* 2: 286
Van der Waals, H. G. y C. O. Roelofs (1930). «Optische Scheinbewegung», *Zeitschrift für Psychologie und Physiologie des Sinnesorgane* 114: 241-248, 115 (1931), págs. 91-190.
Van Essen, D. C. (1979). «Visual areas of the mammalian cerebral cortex», *Annual Review of Neuroscience* 2: 227-263.
Van Gulick, R. (1988). «Consciousness, intrinsic intentionality, and self-understanding machines», en Marcel, A. J. y E. Bisiach (comps.), *Consciousness in Contemporary Science,* Oxford, Clarendon Press, págs. 78-100.
Van Gulick, R. (1989). «What difference does consciousness make?», *Philosophical Topics* 17: 211-230.
Van Gulick, R. (1990). «Understanding the phenomenal mind: Are we all just armadillos?» Manuscrito. Presentado en el congreso «The Phenomenal Mind - How Is It Possible and Why Is It Necessary?» Zentrum für Interdisziplinäre Forschung, Bielefeld, Alemania, 14-17 de mayo.
Van Tujil, H. F. J. M. (1975). «A new visual illusion: Neonlike color spreading and complementary color induction between subjective contours», *Acta Psychologica* 39: 441-445.
Vendler, Z. (1972). *Res Cogitans,* Ithaca, NY, Cornell University Press.
Vendler, Z. (1984). *The Matter of Minds,* Oxford, Clarendon Press.
Von der Malsburg, C. (1985). «Nervous structures with dynamical links», *Berichte der Bunsen-Gesellschaft für Physikalische Chemie* 89: 703-710.
Von Uexküll, J. (1909). *Umwelt un Innenwelt der Tiere,* Berlín, Jena.
Vosberg, R., N. Fraser y J. Guehl (1960). «Imagery sequence in sensory deprivation», *Archives of General Psychiatry* 2: 356-357.
Walton, K. (1973). «Pictures and make believe», *Philosophical Review* 82: 238-319.
Walton, K. (1978). «Fearing fiction», *Journal of Philosophy* 75: 6-27.
Warren, R. M. (1970). «Perceptual restoration of missing speech sounds», *Science* 167: 392-393.
Wasserman, G. S. (1985). «Neural/mental chronometry and chronotheology», *Behavioral and Brain Sciences* 8: 556-557.
Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications,* Oxford, Oxford University Press.

Weiskrantz, L. (1988). «Some contributions of neuropsychology of vision and memory to the problem of consciousness», en Marcel, A. y E. Bisiach (comps.), *Consciousness in Contemporary Science,* Nueva York, Oxford University Press, págs. 183-199.
Weiskrantz, L. (1989). Mesa redonda sobre la conciencia. European Brain and Behavior Society, Turín, septiembre de 1989.
Weiskrantz, L. (1990). «Outlooks for blindsight: Explicit methodologies for implicit processes (The Ferrier Lecture)», *Proceedings of the Royal Society London* B 239: 247-278.
Welch, R. B. (1978). *Perceptual Modification: Adapting to Altered Sensory Environments,* Nueva York, Academic Press.
Wertheimer, M. (1912). «Experimentelle Studien über das Sehen von Bewegung», *Zeitschrift für Psychologie* 61: 161-265.
White, S. L. (1986). «The curse of qualia», *Synthese* 68: 333-368.
Whiten, A. y R. Byrne (1988). «Toward the next generation in data quality: a new survey of primate tactical deception», *Behavioral and Brain Sciences* 11: 267-273.
Wiener, N. (1948). *Cybernetics: or Control and Communication in the Animal and the Machine,* Cambridge, Technology Press. [Trad. cast. de F. Martín: *Cibernética o el control y comunicación en animales y máquinas.* Barcelona, Tusquets, 1985.]
Wilkes, K. V. (1988). *Real People,* Oxford, Oxford University Press.
Wilsson, L. (1974). «Observations and experiments on the ethology of the European beaver», *Viltrevy, Swedish Wildlife* 8: 115-266.
Winograd, T. (1972). *Understanding Natural Language,* Nueva York, Academic Press.
Wittgenstein, L. (1953). *Philosophische Untersuchungen,* Oxford, Blackwell. [Trad. cast. de Alfonso García Suárez y Ulises Moulines: *Investigaciones Filosóficas,* Barcelona, Crítica, 1988.]
Wolfe, J. (1990). «Three aspects of the parallel guidance of visual attention», en *Proceedings of the Cognitive Science Society,* Hillsdale, NJ, Erlbaum, págs. 1.048-1.049.
Yonas, A. (1981). «Infants' responses to optical information for collision», en Aslin, R. N., J. R. Alberts y M. R. Peterson (comps.), *Development of Perception: Psychobiological perspectives vol. 2: The Visual System,* Nueva York, Academic Press.
Young, J. Z. (1965a). «The organization of a memory system», *Proceedings of the Royal Society of London [Biology]* 163: 285-320.
Young, J. Z. (1965b). *A Model of the Brain,* Oxford, Clarendon.
Young, R. Z. (1979). «Learning as a process of selection», *Journal of the Royal Society of Medicine* 72: 801-804.
Zajonc, R. y H. Markus (1984). «Affect and cognition: The hard interface», en Izard, C., J. Kagan y R. Zajonc (comps.), *Emotion, Cognition and Behavior,* Cambridge, Cambridge University Press, págs. 73-102.
Zeki, S. M. y S. Shipp (1988). «The functional logic of cortical connections», *Nature* 335: 311-317.
Zihl, J. (1980) «"Blindsight": Improvement of visually guided eye movements by systematic practice in patients with cerebral blindness», *Neuropsychologica* 18: 71-77.
Zihl, J. (1981). «Recovery of visual functions in patients with cerebral blindness», *Experimental Brain Research* 44: 159-169.

EL AUTOR QUISIERA AGRADECER EL PERMISO PARA REPRODUCIR LAS SIGUIENTES FIGURAS:

Figura 2.3. © 1969 Harvey Comics Entertainment, Inc.
Figura 2.4. © 1975 Sidney Harris - *American Scientist magazine.*
Figura 4.1. De *Shakey the Computer.* Nils Nilsson. Copyright © 1984. SRI International.
Figura 4.3. De *The Thinking Computer: Mind Inside Matter.* Bertram Raphael. Copyright © 1976 W. H. Freeman and Company.
Figura 5.1. De *A Brief History of Time.* Copyright © 1988 Stephen W. Hawking. Bantam, Doubleday, Dell Publishing Group, Inc.
Figuras 5.5. y 5.6. De «Mathematical Games» de Martin Gardner. *Scientific American,* diciembre de 1981.
Figura 5.7. De *Foundations of Cyclopean Vision.* Bela Julesz. Copyright © 1971 University of Chicago Press.
Figura 7.4. Basada en MacroComputer, una simulación por ordenador interactiva de un ordenador, desarrollada por Steve Barney en el Curricular Software Studio, Tufts University.
Figura 8.1. De *Speaking: From Intention of Articulation.* Willem J. M. Levelt. Copyright © 1989 Massachusetts Institute of Technology.
Figura 9.1. De *The Architecture of Cognition* de John R. Anderson, Cambridge, MA: Harvard University Press, Copyright © 1983 The President and Fellows of Harvard College.
Figura 10.1. De *Brainstorms* de Daniel Dennett. Copyright © 1978 Bradford Books, Publishers. Publicado por MIT Press.
Figura 10.7. Dibujo de Gahan Wilson © 1990 *The New Yorker Magazine,* Inc.
Figura 11.3. De *Brain Mechanisms in Sensory Substitution.* Paul Bach-y-Rita. Copyright © 1972 Academic Press, Inc.
Figura 11.4. De *Nature,* vol. 221, págs. 963-964. Copyright © 1969 Macmillan Magazines Ltd.

ÍNDICE ANALÍTICO

Abelson, R., 271, 449, 497
aborto, 466
Abrahamson, A., 282, 479
absolutismo, 431, 465
abstracción, el yo como, 424-427, 429, 431, 437
abstracta, 103, 378
abuso sexual, 430
acceso, 243
— el de Shakey frente al nuestro, 105, 323
— inmediato, 331
— limitado, 386
— menos que óptimo, 209
— privilegiado, 76, 81, 136, 259-260
acción intencional, 43, 265
acción moral, 36-37
aceleración, 47
acte gratuit, 177
activación, umbral de, 179
actos frente a acontecimientos, 44
ACT, 278-283, 285, 293
acumulador, 226 249, 250
adaptación a la visión invertida, 404-406, 408-409
«Adecuación neuronal», 168, 171, 173-174
adivinar cuándo hay que adivinar, 343-344
ADN, 37
afasia, 262
afecto, 57, 61-62
— y color, 393-400, 405-406
agente racional, 89
agujero negro, 48, 116n., 185
— como entidad teórica, 84
ajedrez, 312-313
Akins, K., 193, 264n., 388, 394, 453-455
alegoría, 96, 106
algorítmico, nivel, 289
algoritmo, 238
algoritmo de comprensión, 360
Allison, M., 262
Allman, J., 148, 479
Allport, A., 273, 288n., 289, 479
alma, 44, 440, 463-464
— inmortalidad del, 378-379
— teoría de los centros de gravidad, 378
alucinación, 15-29, 53, 132, 252
— auditiva, 264n.
— de la memoria, 132
ambigüedad de alcance, 257, 260
amígdala, 53
amnesia, 263, 407
amor, 34-36
análisis por síntesis, 23-24
analizador cortical, 368
ancho de banda, 18, 253n.
andares de araña, 474
Anderson, J., 272, 278-280, 479
Anderson, Jane, 14
Andler, D., 14
anestesia
— fingida, 52
— local, 168
animado, 44
anosognosia, 91, 263, 366, 368-369
Anscombe, G. E. M., 328n., 344, 479
anteojos que invierten la imagen, 404, 408-410
anticipación, 191-193, 195, 201
antimateria, 48
antropología, 94-98, 108, 268-272
aparato, la mente como, 293-294
aparecer repentinamente, 192, 352, 408
apariencia, 147, 332, 375, 377, 386, 420
— frente a realidad, 130-131, 144-145
— real, 147, 328, 374-375
apetito epistémico, 27, 194, 336, 366-367, 369-370, 393
apreciación, 42, 44
aprehensión, interna, 317
— lapso de, 154
aprendizaje, 206-207, 284
— como rediseño, 196
— por imitación, 234
— por instrucción explícita, 234
apuntar, 344-345
araña, 425-426, 428, 452
archivo, carpeta de, 360
Aristóteles, 206n.
Arnauld, A., 333, 479
arquitectura de pizarra, 278-279
articulación, nivel adicional de, 317
ascidia, 191
ASCII, código, 249
asco, 425
asociación, 238, 399, 453
atención, 238, 396
— captación de la, 203
ausencia de representación vs. representación de la ausencia, 374
ausencia (durante crisis epilépticas), 366-367
auto-
— conocimiento, 459-461
— control, 235, 290-292, 427-428, 437-439, 459
— estimulación, 208-209, 222, 232, 238, 289, 314-315, 361
— exploración, 223, 241-242, 268, 313
— fenomenología, 109
— indicio, 443
— interpretación, 259, 437
— manipulación, 222, 291, 294, 308, 339
— observación, 328
— reconocimiento, 437-439
— representación, 323, 428-429, 439-440

— sondeo, 183
— supervisión, 235, 263, 332, 341
— transparencia de la mente, 314
autómata, 44, 85, 269, 334, 337, 427
automatista, 269
autor, 376-378, 428-429
— aparente, 94
— autoridad del, 94, 109, 258, 317
— de las grabaciones, 242
— de los actos de habla, 91, 243-244
— intenciones del, 94, 96
aviario de Platón, 236, 239, 279, 284, 293, 314
ayuda, su papel en la visión ciega, 339-143

Baars, B., 271-272, 284, 292, 356, 368, 479
Bach, J., 398-399, 453
Bach-y-Rita, P., 14, 350-353, 479
balístico/-a,
— acción, 159
— acontecimientos no referibles, 348
— carácter de los movimientos sacádicos, 372
Ballard, D., 202, 479
banda, pequeña, en el cerebro, 59
banda sonora, doblaje, 126, 166
Barnum, P., 257n.
Baudot, J., 253n.
Beattie, J., 14
Bechtel, W., 282, 479
Beethoven, L., 43, 59-60, 214
béisbol, 261
Bennett, J., 207, 391, 480
Berkeley, G., 67, 79
biblioteca, como analogía de la visión, 371
Bick, P., 264, 480
Bieri, P., 295, 480
biorrealimentación (*biofeedback*), 343
Birnbaum, L., 255-257, 480
Bisiach, E., 270, 369, 480
bit (dígito binario), 231, 353n.
— definidos, 226
— mapa de, 305-306, 309-310, 360, 364-365
Block, N., 14, 147n., 410, 447, 451, 480
Bloom, P., 213, 495
Boghossian, P., 470, 480
Booth, W., 85n., 480
Borges, J., 480
Borgia, G., 426, 480
Boyle, R., 382
Brainstorm (película), 401
Braitenberg, V., 185, 192, 481
Breitmeyer, B., 155, 481
bridge, 313
Broad, C., 413-414, 481
Bronowski, J., 222
Brooks, B. A., 372, 481
bucle infinito, 280
bucles temporales, en la secuencia subjetiva, 150, 168
buffer
— de fotogramas, 360
— de memoria, 231, 233
— visual, 300
burocracia de homúnculos, 250, 252, 253, 257, 264, 275, 440
Byrne, R., 208n., 457, 481, 501

cadáveres, tratamiento de los, 453-464
CADBLIND, sistema, 305-306, 308-311, 325, 375, 386, 415-416, 451
CAD, sistema de, 300-301, 303-306, 308, 357
caja negra de la conciencia, 185, 241
cajología, 284n., 369n.
Calvanio, R., 481
Calvin, W., 14, 128, 159, 177, 197n., 203, 238, 272, 288, 481
Campion, J., 339n., 481
campo electromagnético, 116n.
Camus, A., 33-34, 481
Canli, T., 14
cantata, 398-399, 453
capacidad de discriminación burda, 192
Carnegie, D., 315
Carruthers, P., 339, 481
cartesiano (*véase también* Descartes, R.)
— cuello de botella, 119
— interaccionismo, 46
— materialismo, 121, 133, 153, 158, 270, 332-333, 355
— Teatro, 29, 51, 121, 125, 127-129, 140, 141, 145, 147, 150-153, 156-157, 171, 178-180, 183, 185, 241, 243, 244, 267
Casa Blanca (*véase también* Despacho Oval), 178
Casper el Fantasma Bueno, 47
Castaneda, C., 19, 481
Castañeda, H.-N., 481
castor, 203, 425-426, 428
Catalina la Grande, 21
categorial, error, 61
causa y efecto, el orden de, 165-166
causalidad
— mecánica, 170
— no observable, 147
ceguera, 67, 285, 350-353
— histérica, 90, 338-389,
— negación de la, *véase* síndrome de Anton
— psicosomática, 338
Centro de Gravedad Narrativa, 421, 429, 437, 439-440, 450, 464, 459
centro
— de gravedad, 108, 115, 378, 423-424, 429
— del cerebro (donde se supone que se produce la experiencia consciente), 51, 118, 121, 124-125, 178, 179, 250, 274
cerebro
— como máquina, 43
— como mente, 28, 44
— electrónico gigante, 228, 231
— en un tarro, 15-16, 19, 28, 452
— perla cerebral (*véase también* perla mental), 434, 440, 379
— ¿piensa?, 41
cerveza, como gusto adquirido, 407-408
Chambers, D., 304, 308, 496
Changeux, J.-P., 197n., 481
Chaplin, Greta y Freda, 432-433
Cheney, D., 207, 455, 457, 481
Cherniak, C., 293n., 481
chimpancé, 203, 207
— auto-reconocimiento en el, 438
Chomsky, N., 203, 245, 314
Churchland, P. M., 410, 482
Churchland, P. S., 168, 173-176, 482

ciclo de buscar y ejecutar, 227
ciencia cognitiva, 29, 51, 170, 249, 269-270, 274, 412
científicos, 49
cinéma vérité, 21
cíngulo anterior, 121
circuitos para la coordinación de ojos y manos, 201
claqueta, 166, 172
clarividencia, 140
clase natural, 393n.
CLEF, el acrónimo, 202
código de colores (*véase también* color), 361, 373, 415
cogito ergo sum, 80
Cohen, L. D., 73, 482
Cole, D., 406, 482
Cole, P., 421n.
Collins, G., 255, 256, 480
Collins, W., 373, 495
color, 355, 383-397, 465, 470
— ceguera al, 390
— células que se interesan por el, 285
— código de, 308, 388-389, 396
— código númerico para el, 357-358, 415, 363-364
— discriminación por parte del, 148, 162
— experiencia del, 39, 40
— fenómeno phi del, 128-129
— sólido de, 361
— visión en, 42, 66-67, 193, 361-365, 410-412
comillas, 355, 357
comisurotomía, 423, 436
competición en el cerebro (*véase también* evolución; redes Quien Gana se lo Queda Todo), 252, 288-289, 348, 368, 452
comprensión, 67-70, 293, 450-451
— del lenguaje, 82
— ¿dónde se produce?, 333, 446
— fenomenología de la, 68-70
comunicación, 253n.
— en el cerebro, 210-211, 288, 294, 306-308, 314, 326, 328, 332, 337
— verbal, su importancia en los experimentos, 86
Conan Doyle, A., 94, 419
concentración, 291
concepción clara y distinta, 295
concepción frente a imaginación, 295
conceptos, fenómenos que dependen de sus, 36
Conceptualizador, 247, 248, 250, 252, 254, 255n., 264
conciencia, *passim* (*véase también* flujo de la conciencia), 180
— como subsistema, 180
— conceptos de, 35
— como modo de acción y no como subsistema, 180
— de los animales, 29, 74, 85-86, 453-463
— discontinuidad de la, 366, 372, 433
— estructura oculta de la, 445
— unidad de la, 122
conductismo, 52, 83, 197, 472
— desnudo, 83, 170
conejo cutáneo, 156-157, 170
conexión reforzada, como microhábitos, 231-232
conexionismo, 189n., 253, 282-283
confabulación, 106, 263
conocimiento del mundo, 247, 250-251
constructo lógico, 95
contenido (Juicio) (*véase también* intencionalidad), 141, 365, 376, 467
— cualitativo, 384
— de la experiencia, 81, 156
— enviado por las neuronas, 276
— fijación del, 376, 443
— indeterminado, 260
— individuación de estados por su, 331
— predeterminado, 250
— vehículos de, 145
contorno subjetivo, 362
contraintuitivo, 49, 112, 424, 441
control (*véase también* auto-control), 178, 183, 193, 201-202
— flujo de, 250
— usurpado y no delegado, 254
Cooper, L., 107, 365n.
corpus callosum, 434
córtex occipital, 334
— prefrontal, 121, 289
— somatosensorial, 168-169
Cowey, A., 337, 499
CPU (unidad central de procesamiento; central processing unit), 232, 249
Crane, H., 82n., 482
creatividad, 258, 311
— en el uso del lenguaje, 475
creencias, 89-90, 207n., 323
— entorno de, 463-464,
— expresión de, 91, 97, 106, 110, 144, 146, 319, 330
— expresión equivocada de, 97
— fijación, 274
Crick, F., 180, 269-270, 287, 482
cualidades encantadoras y sospechosas, 391-392
cualidades, primarias y secundarias, 382-383, 384, 387, 390-394
Cuartel General, 120, 159, 178, 188, 190, 428
cuello de botella de von Neumann, 227, 249
Curva Laffer, 123
Curva Neo-Laffer, 124, 406

dado, 145
Dahlbom, B., 14
Damasio, A., 82, 482
Damasio, H., 82, 482
Danchin, A., 197n., 482
Darwin, C., 186, 197n., 211-212, 222, 431, 470
datos sensoriales, 333-334
Davis, W., 72n., 482
Dawkins, M., 460-462, 482
Dawkins, R., 187, 208, 213-216, 217, 219-221, 426, 432, 456, 482-483
«de abajo arriba» vs. «de arriba abajo», 12
de Kooning, W., 396
de Sousa, R., 435, 483
Dehaene, S., 197n., 482
deliberación, 37
demonio (*véase también* homúnculo), 202, 252, 275-278, 286, 290, 296, 312, 319, 370, 421, 450
— la producción como, 280
— malvado, 15, 18, 22
Dennett, D. C., 48n., 51, 56, 84n., 89, 90n., 108, 130n., 145n., 151, 154n., 158, 168, 189
denotabilidad (*véase también* intencionalidad), 344-346, 376-377, 383
Derrida, J., 421

desarollo, como rediseño, 196
Descartes, R. (*véase también* cartesiano), 13, 15, 18, 20, 41, 42, 45-47, 49n., 53, 55n., 79-80, 82, 99, 117-118, 119, 121
«desde dentro», 53
deseo, 90, 207n., 461
desmitificación, 34, 37
Despacho Oval en el cerebro, 118, 120, 426-427, 440
detección de colisiones, 195
deus ex machina, 37
dibujar, 64
dicromática, visión, en el ciervo, 454
dinero, 36
disciplina
— del ajedrez, 313
— del lenguaje, 314
discriminación por parte del cerebro, 148-149
— del orden temporal, 167
— distribuida, 311
— sólo tiene que producirse una vez, 141, 142, 167
discriminativo, estado, 384
diseño de artefactos frente a diseño de la naturaleza, 189
disfrute, 37, 42-43, 74, 355, 366-367, 395-400, 415, 420-421
disposición, 384-385, 397-401
— mental, 251-252, 254
— propiedad disposicional, 387, 391, 470
— reactiva, 400, 402-406, 407-410, 415
dispositivo dependiente del contenido, 166
distinción entre hemisferio derecho e izquierdo, 195, 211, 228, 434
diversión, 74, 461
— como variable intermedia, 77
dolor (*véase también* sufrimiento), 36, 41, 71-73, 75, 211, 331, 397, 460-465, 466
— ¿es real?, 470
— proyectado, 142
— psicosomático, 338
— referido, 321
— soñado, 72
dolor, por qué lo provocan los dolores, 72-73
«¿dónde va a parar todo?», 51, 120-121, 147-148, 179, 310-311, 378
Dorfman, A., 129, 145
dorsales y ventrales, sistemas cerebrales, 194
Dreyfus, H., 283, 292-293, 485
Dreyfus, S., 283
drogas, sus efectos en las alucinaciones, 25
dualismo, 39, 45-53, 119, 140, 147, 170-171, 173n., 269, 355, 369-370, 450

Eccles, J., 41, 45n., 49n., 168-169, 173, 485, 495
ecolocalización, 452-458
Eco, U., 259, 485
ectoplasma, 47
ecuaciones de Lorentz, 116, 116n.
ecuaciones de Maxwell, 116n.
ecuador, 108
Edelman, G., 189n., 197n., 281, 326n., 367, 377, 470, 486
edición en el cerebro, 125, 134, 148, 157, 167
— cambios editoriales frente a erratas, 139, 260
— sala de, 134-136, 166, 173
Efecto Baldwin, 197-200, 203, 210, 221
efecto de reconstrucción de los fonemas, 355
efecto del lector de pruebas, 356
Efecto Doppler, 115, 456-457
Efecto Liebmann, 82n., 476, 492
Efecto McGurk, 125
efectos secundarios inesperados, 189
Efron, R., 161, 486
ego, 29, 227
egoico, 187, 427, 458
egoísmo, variedades de, 188
élan vital, 37
Eldredge, N., 432, 486
elección forzada, 338-341, 343, 347n., 475
electromiograma, 177
eliminación de líneas ocultas, 302
empirismo inglés, 67, 71, 79
encapsulamiento, 254, 273
encontrar alimento, 203
energía física, 47
enfermedad, 36
«enlatado», lenguaje, 243, 250, 322
ensamblador, lenguaje, 315
ensayo, 238, 291-292, 438
epifenomenismo, 83, 409, 412-414, 471
epífisis (*véase también* glándula pineal), 46
epilepsia, 366-367, 433-435
epitelio, 42, 58
epojé, 56
Ericsson, K., 87, 486
error, 259-260, 329-332, 425
errores lingüísticos, 246
escarabajo en la caja, 402
criterios comportamentales, 417
esconde el dedal, 344-347
escotoma, 335-337, 341, 344, 364n., 366
escritura, su importancia en la formación de la conciencia, 234
esencialismo, 431
espacio
— cualitativo, 392-394, 405, 407, 415
— de problemas, 280
— del representante frente al del representado, 157-158
— fenoménico, 144, 317, 367
— lógico, 144, 145
— personal, 347n.
— semántico, 260, 261n.
— sólido de colores como ejemplo, 361
— visuo-motor, 409
espacio de trabajo, 226, 249, 269, 277-278, 279-281, 283-284, 294-295
— global, 271, 284
espíritu, 9, 429-430
— en la máquina, 45, 273-274, 424, 434-435
esquizofrenia, 264n.
estado intencional de orden superior, 319-321, 326
estaliniano (*véase* orwelliano frente a estaliniano)
— el modelo de Libet, 171, 173-174, 177-178
estereograma de puntos aleatorios, 125, 166
estrategia del diseño, 289n.
éter, 48
etiqueta, 306, 363-366
etología cognitiva, 457
eutanasia, 465
Evans, G., 272, 486
evolución, 33, 73-75, 185-239, 250, 255, 290, 307, 395, 412-414, 420, 431, 443

— cultural (*véase también* mema), 206
— de la visión en color, 388-394
— del dolor, 72
— de los yos, 423-426
— velocidad de, 199-200
— y sufrimiento, 462
Ewert, J.-P., 201, 486
existencialismo, 177
expansión del color de neón, 361, 374
experimentación con animales, 465
experimental, método, 337
experimento mental (*véase también* bomba de intuiciones), 16, 28-29, 60, 335, 378, 402-404, 408-410, 411, 447-448, 452
experimentos, diseño de, 86-87
explicar frente a eliminar, 465-466
exploración (*véase también* auto-exploración), 461
explosión combinatoria, 17, 27
expresión (*véase también* creencias, expresión de), 183, 254, 260, 318

Fallon, A., 425
fantasma, cuerpo, extremidad, 17, 20
FAP (patrones de acción fijos), 273
Faraday, M., 56
Farah, M., 107, 308n., 486, 491
Faucher, L., 14
fechado, 160-161, 166
Fehrer, E., 155, 486
Feldman, J., 202, 479
Fenhomo, 94-98, 107, 108, 144
fenol-thio-urea, 391
fenoménico
— campo, 29, 63, 81
— espacio, 143, 317, 367
— propiedad, 334, 349, 351-352, 384, 394
Fenomenología (escuela filosófica), 55-56, 292
fenomenología, 55-56, 59, 66, 71, 76-80, 96, 108, 147, 171, 265, 269, 309, 361, 433
— como la conducta de una caja negra, 185
— de la comprensión, 68
— de la risa, 75-76
— paradójica, 137
— pura e impura, 82
— «real», 376-377
— visual, 66, 67
fenotipo, 195
— evolución en el, 196-197
— extendido, 426
Feynmann, R., 307-308
ficción, 376-377, 419, 449
— de un teórico, 93-95, 141, 171, 378
— gramatical, 472
— interpretación de la, 106
— literaria frente a la ficción de un teórico, 107-111
— tratada como un hecho, 96-97
— y mundos ficticios, 91-94
figmento, 357, 361-362, 364, 366, 370, 376-378, 381-383, 386, 445-446, 466
figura y fondo, 346
fijación postnatal del diseño, 196, 213
filósofos, 15, 29, 34, 49, 52, 76, 85, 99, 128, 150, 253n., 268, 292, 296, 334, 337, 381-382, 401, 410, 412, 416, 431, 445, 446-447, 452, 467-472
Finke, R., 308n., 486
fisicalismo (*véase también* materialismo), 411
física, 48-49, 170
— revolución en la, 384
Flanagan, O., 14, 287, 486
«Flight Simulator», 302n.
Flohr, H., 14, 287, 486
flujo de la conciencia, 56, 80, 127, 148-150, 151, 158, 179-180, 203, 228, 238, 249, 267, 271-272, 267, 418, 466
Fodor, J., 206n., 254, 273, 284n., 286n., 293, 486
— como cripto-cartesiano, 274
fondo de la (segundo plano) conciencia, 347, 365, 373, 419, 477
— como contexto, 291-292
formación
— del cerebro, 232
— del oído, 348
— del paladar, 407
formación reticular, 121, 287
formulador, 248, 250-251, 253-254, 255n.
Forster, E., 207, 259
fóvea (*véase también* visión parafoveal), 65-66, 366-367
Fox, I., 381, 413, 486
Fraser, N., 25, 500
French, R., 323, 487
Freud, S., 25-26, 97, 252, 256-258, 261n., 327, 487
Freyd, J., 300n., 487
función, 187
— de la conciencia, 235, 289-291
— de la visión, 336-337
— múltiple, 189-190, 285-287, 290-291
funcional frente a epifenoménico, 412-415
funcionalismo, 42-43, 284n., 337, 397, 468-470
— de máquina de Turing, 469
— homuncular, 275
— teleofuncionalismo, 469
Fusillo, M., 369, 487
Fuster, J., 288, 487
futuro, producir, 158, 190-192, 201, 291

gallinas, sus preferencias, 462
Gardner, H., 334n., 487
Gardner, M., 124, 487
gato, lo que se siente al ser, 452
Gazzaniga, M., 207, 211, 273, 434, 487
Gelade, G., 347n., 500
Geldard, F., 156-157, 487
generación y verificación, 23-25, 206
genotipos, 195
Gentilucci, M., 201, 496
gen, como entidad teórica, 84
Gert, B., 487
Geschwind, N., 369, 487
Gide, A., 177
Gilbert, W., 56
glándula pineal, 46, 53, 118-122, 183, 271
GOFAI (Good Old Fashioned AI; Buena y Vieja I), 282
Goodman, N., 128, 134, 140-142, 145
Goody, J., 272, 487
gorila, 203, 207
Gould, S., 432, 487
Gouras, P., 487
grabación y reproducción en alta fidelidad, 59-60, 62-62

gramática, como fuente de disciplina, 314-315
gramatical, estructura, 63
Gran Encefalización, 203
Graubard, S., 282, 487
gravedad (*véase también* Centro de Gravedad Narrativa), 196
— centro de, 108, 115, 379, 423-424, 429
— efectos sobre la luz, 115
Gregory, R., 364n., 476, 495
gremlin, 414-415, 417
Grey Walter, W., 180-181, 475, 487
Grice, H., 207n., 487
Griffin, D., 457
Grupos neuronales, 189n.
Guehl, J., 25, 500
Guerra de las galaxias, La, 444
guión, 271, 449
guitarra, 60-61

Habitación China, experimento mental, 334, 447-450
hábito, 453
— de mente, 194, 234-235, 237, 242, 268, 277
habla
— actos de, 88-91, 134, 183, 242, 247-249, 251-253, 260-262, 263-264, 314-315, 326, 328, 376, 467
— comprimida, 158
— creativa, 258
— percepción del, 158
— producción del, 245, 251, 254, 262, 328
— sonidos del, 62, 63, 86-87
hablar con uno mismo, 208-210, 235, 237, 289, 312, 418
— frente a subsistemas hablando los unos con los otros, 329
— ¿para qué sirve?, 314
hacer un seguimiento, 307
Hacking, I., 459, 488
Hal (en *2001*), 444
Hamlet, 26
Hampl, P., 258, 488
Handford, M., 291, 488
Hardin, C., 52n., 82n., 356-357, 361n., 387-388, 403n., 453, 488
Harman, G., 381, 410, 413, 488
Harnad, S., 176n., 177, 234, 322, 324, 488
Haugeland, J., 282, 488
Hayes, P., 271, 488
Hayes-Roth, B., 278, 488
Hebb, D., 189n., 222, 470, 488
Heeschen, W., 262n.
hemianopia, 335
hemi-negligencia, 357-369
heterofenomenología, 84-92, 96, 222, 244, 277, 337, 371, 376, 407, 429, 446, 464, 468
— de los animales, 455-465
— neutralidad de, 84, 107-222
heterofenomenológico, mundo, 97, 105-109, 111, 134, 138, 140, 170, 242-243, 260-261, 308, 311, 332, 338, 377, 400, 418, 421, 458-461, 469-470
heurística, 293
hilaridad, 75-76
— comparada con el horror, 461
Hilbert, D., 14
Hilbert, D. R., 387, 488
Hintikka, J., 468, 488
Hinton, G., 199, 488
hipnogógica, ensoñación, 313
hipnótica, sugestión, 72
hipótesis, verificación de, 23-25
Hobbes, T., 142, 489
Hoffman, R., 264n., 489
Hofstadter, D., 14, 235, 252, 322-324, 421n., 447n., 453, 489
Hogeweg, A., 189
Holcomb, P., 14
holismo, 283
Holland, J., 197n., 489
Holmes, Sherlock, 92, 94, 144, 419-420
Holtzmann, J., 107, 491
Holyoak, K., 197n., 489
homínido, 203-204, 207-209, 210, 228
homo ex machina, 101
Homo sapiens, 203-204, 207, 215, 222, 274, 288, 426
homogeneidad de propiedades, 382-384
homúnculo, 26, 27, 99, 103, 106, 190, 242, 263, 272-276, 336, 366, 368, 440, 466
— el centinela parafoveal como, 370-373
— guardia de tráfico, 280
Honderich, T., 168, 489
Hora de Greenwich, 176
horizonte de simultaneidad, 120
Howell, R., 91, 489
Hu-Min, J., 447n.
Hubel, D., 52, 148, 492
Hughlings Jackson, H., 207, 489
Hume, D., 49n., 67, 79, 195, 214, 423-424, 489
Humphrey, N., 14, 44, 63, 191, 235, 388, 396, 430n., 434, 439, 489
Hundert, E., 335, 490
Husserl, E., 56
Huxley, T., 413, 415-416, 490

I + D, 221
IA (Inteligencia Artificial), 51, 69, 104, 197, 189, 229, 231, 235, 255, 269, 271, 275, 277, 282, 293, 451, 463
— críticos de, 283
— fuerte, 447, 451, 465
IBM-PC, 232-233, 302n.
idea
— concepto de Descartes y de Hume, 49n.
— de rojo, 383, 384-386
— simple, 214
idealización, 468-469
identidad, teoría de la, 169, 172, 468-470
idoneidad, 214, 216, 218
— de los memas, 235
— dependiente de la frecuencia, 220
ilusión
— benigna, 376
— de la conciencia, 443-445, 446-447, 466
— del usuario, 229, 233, 324-325
— y color, 386-387
Ilusión de la Inmanencia, 371
ilusionista (*véase también* truco de magia), 20, 22, 25
imagen, 67-69
— en la cabeza, 64-66, 69-70, 81, 245, 311
— mental, 98, 106-107, 143-145, 177, 243, 268, 311-312, 360, 365n.
— real, 103-105

— realidad de las imágenes mentales, 468-470
— rotación de, 299-311, 365n.
imágenes como objetos de nuestra percepción visual, 63
imaginación, 39, 40, 61, 71-72, 377, 408-409, 443, 459-460, 410-413, 443-447, 449-450, 461
— bloqueo de la, 399
— cómo se construye, 67
— ensanchamiento de la, 7, 98, 267, 295
— falta de, 29, 63-64, 294-296, 316, 381
— límites de la, 452
— muletas para la, 304, 332, 450-451
— su papel en la diversión y el sufrimiento, 461
impasse, 280
impenetrabilidad cognitiva, 274
Imperio Británico, 160, 176, 183
importancia, 43, 53, 187, 459-466
Impresionismo, 56, 66
inconsciente, 338
— acción iniciada inconscientemente, 174, 177
— calumnia, 96-97
— cognición, 294-295
— conducción, 151
— control, 340
— objetivos, 256-257
— pensamiento, 321-322
— pensamientos de orden superior, 320-322, 324
— procesos, 37-38
— producción del lenguaje, comprensión, 109-110
— protocolos automatizados, 340-341
incorregibilidad, 76-77, 80, 351
indeterminación de los mundos ficticios, 93
inefalibilidad, 60-62, 108-110, 385, 393-395, 399
infalibilidad
— de la introspección, 80
— papal, 95, 109
información
— complejidad, 388, 412
— medición (*véase también* ancho de banda, nivel de baudios; bit), 18
— nuestra capacidad de utilizar, 27
— paso desde fuera hacia dentro,67
— poca, en la visión ciega, 337-339, 342, 353-370
— procesamiento, en el murciélago, 455
— su transmisión y su procesamiento en el cerebro, 37-39, 158, 161, 353n., 344
informar, neutralmente, 191
inforvívoro, 194
ingeniería, 86
— genética, 222
— inversa, 159, 160, 290n.
Iniciativa de Defensa Estratégica, 164
inmortalidad, 441
inmunología, 206n.
inmunológico, sistema, 188
innato
— conocimiento, 279-280
— mecanismos del lenguaje, 203-204, 213, 233
— mecanismos y estructuras, 190-191, 195, 233, 268, 272-273, 276, 400
— predisposiciones en el espacio cualitativo, 404-405
— tendencia, 44
instrumentalismo, 468-470
Inteligencia Artificial. *Véase* IA como si, 377
intención
— comunicativa, 207, 247, 251-258, 259, 261, 263, 329-332
— consciente, 175-178
— del autor, 96, 258-259
intencional
— acción, 43, 181
— actitud, 88-90, 207n., 289n., 467-468
— objeto, 95, 107-108, 110-111, 144-145
intencionalidad, 88-92, 205, 344-347, 467-468
— definición, 344
— intrínseca, 292
interaccionismo, 46, 53
«interesadas por», neuronas, 276, 285, 336, 368, 404
interpretación, 91, 109-110
— auto-interpretación, 259-260
— de la ficción, 96
— de los estados cerebrales por el mismo cerebro, 326n.
— del habla, 88-90, 86-89, 242
— de un texto, 91-93, 258-260, 376-378
— por parte del cerebro, 125, 178-179
introspección, 29, 67, 76-79, 106-107, 176n., 236, 244, 269, 232
— como algo que comporta un pensamiento de tercer orden, 320-322
— como teorización, 79-81, 106-107
— de la máquina von neumanniana, 227-229
— del yo, 423
— limitaciones de la, 364-365
introspeccionista, movimiento en psicología, 55-56, 72, 83
intuición, 373-374, 381, 405
— bomba de, 410-412, 413, 448-449, 453
invertido, qualia o espectro, 400-410, 469
invisibilidad de la estructura funcional, 222, 233
IRM (imagen por resonancia magnética), 336
IRM (Innate Releasing Mechanisms; Mecanismos Innatos de Desencadenamiento), 273
isotropía, 274, 292-294

Jackendoff, R., 14, 145, 292, 306n., 355-357
Jackson, F., 409, 412, 447, 451, 490
Jacob, F., 190, 490
Jacob, P., 14
Jaikumar, M., 14
James, W., 115, 227, 234, 242, 424
Janlert, L.-E., 307, 490
jardinero, pájaro, 426
Jarrell, R., 455, 490
Jaynes, J., 235, 490
Jefe, 242-243, 247, 265, 434, 431
— en el córtex prefrontal, 288-289
— en el tálamo, 288
Jell-O, caja de gelatina, 388, 393-395, 412
jergafasia, 262
jerga, como expresión de nuevos conceptos, 223-224
Jerison, H., 272, 490
Johnson, D., 421n.
Johnson-Laird, P., 490
Joslin, D., 14
joyceana, máquina, 227, 232, 238-239, 243, 288-295, 458-459
Joyce, J., 225
juego de manos (*véase también* magia), 141, 295, 324, 344

juegos de palabras, 256
juicio, 141, 331, 334, 355, 375-378, 383
— de simultaneidad o de secuencia, 179-181
juicio estético, 71
Juslesz, B., 125, 269, 490
juzgar y parecer, 146-148

Kafka, F., 440
Kant, I., 56, 146n., 153
Keeble, G., 396, 490
Keller, H., 241, 490
Kessel, F., 421n.
Keynes, J., 311, 316
Kim, J., 14
Kinsbourne, M., 154n., 180, 195, 262, 264n., 285, 368, 434, 491
Kirman, B., 73, 491
Kissinger, H., 94
Kitcher, P., 381
kludge, 223, 238
Koch, C., 180, 269-270, 482
Koestler, A., 45, 491
Kohler, I., 404, 491
Kolers, P., 128-129, 133, 135, 140, 142, 154, 491
Kosslyn, S., 107

Lackner, J., 82, 491
Laird, J., 159n., 278, 280, 283, 284n.
lamarckismo, 199, 221
Land, E., 52
Langton, C., 189, 491
lapsus freudiano, 255-257
lapsus, 330
Larkin, J., 491
lateralidad, test de, 147n.
Latto, R., 339n., 481
Lawlor, K., 14
lectura científica del pensamiento, 40
Ledoux, J., 273, 434, 487
Leiber, J., 55n., 315, 453, 491
Leibniz, W., 423, 491
lenguaje, 241, 384
— comprensión, 234
— creatividad del, 311
— del pensamiento, 315-316, 376
— de programación, 249
— efectos sobre la estructura del cerebro y competencia, 213, 220-221, 223, 315
— ensamblador, 315
— evolución del, 203, 208, 213
— máquina, 313, 315
— mecanismos innatos del, 203, 213, 233, 313
— natural, como lenguaje de programación, 315
— percepción frente a producción, 244
— producción (*véase también* producción del habla)
— su papel en la conciencia, 29, 238, 313-314, 457-459
— su papel en la creación de un yo, 426-429
— su papel en la máquina virtual, 427-429
— y proto-lenguaje, 207
lenguaje con segundas, 449
lenguaje de signos, 314, 459
Levelt, W., 14, 245-247, 252-255, 259, 262n., 314, 491
Levine, D., 369, 481
Levy, J., 434, 491
Lewis, D., 91, 410, 491
Ley de Leibniz, 469
Liberman, A., 62, 492
Libet, B., 168-181, 355, 492
libre albedrío, 36-37, 177-178, 181, 440
límite, 424-425, 426-427, 464
— biológico, 188, 190
— de luminancia, 81, 476
— entre yo y el mundo exterior, 122, 188, 190
Lincoln, A., 257, 260
línea del entendimiento (conocimiento), 259
lingüística, 245, 253n., 263
Lisp, 229, 229n., 250, 315
Livinsgtone, M., 148, 492
Llinás, R., 191n.
lo bueno y lo malo, 187-188, 191, 201, 220-221
lo que se siente, 107, 108-109, 110, 201, 398, 435-436, 452-460
Loar, B., 410, 492
lóbulo frontal, 288, 326n.
Locke, J., 67, 79, 214, 382-387, 390, 401, 492
Lockwood, M., 384, 421n., 492
locus ceruleus, 382
Lodge, D., 421, 492
lógica, de las creencias y del conocimiento, 467-468
lugares ligados, 220
luminancia, 411, 418
luz interna, la conciencia como, 459
Luzzatti, C., 369, 480
luz, velocidad de la, 116, 120
Lycan, W., 402, 410, 492

MacDonald, R., 125, 493
Macintosh, 233
MacKay, D., 173n.
Macko, K., 347, 493
Madame Bovary, 92
Madre Naturaleza, 188, 192, 195-196, 274, 286, 293, 390-393n., 425
magia (*véase también* juego de manos), 22, 140, 292-294, 372, 446, 447n.
malo, de pensamiento, 265, 329, 468
mapa retinotópico, 275
máquina
— la persona como, 443-444
— lenguaje, 249, 315-316
— tabla de, 226
máquina catadora de vino, 42-43
máquina de movimiento perpetuo, 47
máquina de von Neumann, 223, 226-232, 238-239, 249-250, 277, 279, 283, 313
Marais, E., 427, 492
Marceau, M., 224, 228
Marcel, A., 13, 270, 339, 341, 492
marco, 275-276
Margolis, H., 204, 233, 236-238, 291, 356, 492
Margulis, L., 425, 492
Marilyn (Monroe), 365-366, 369-371, 375, 407-408, 419, 476
Markus, H., 413, 501
Marler, P., 196, 493
Marr, D., 289-291, 493
Marx, K., 241
Mary, la investigadora del color, 409-413
matasellos, 161, 166, 172
Matelli, M., 201, 496

materialismo (*véase también* materialismo cartesiano), 37, 45, 47-49, 53, 76-77, 119-120, 154, 169-170, 172, 175, 334, 409, 415-418, 440, 450
— eliminativista, 469-470
— semiótico, 421
Maxwell, C., 56
Maynard Smith, J., 186n., 493
McClelland, J., 234, 253, 493
McConnell, J., 14
McCulloch, W., 217n., 493
McGinn, C., 61n., 287, 339n., 445-447, 493
McGlynn, S., 366, 369, 493
McGuinness, E., 148, 479
McGurk, H., 125, 493
McLuhan, M., 396, 493
Meizin, F., 148, 479
Mellor, H., 163, 493
melodías, discriminación de, 475
mema, 213-223, 235, 257, 268, 275, 277, 315, 395, 397, 428
memoria, 251, 254
— a corto plazo, 174
— a largo plazo, 51, 283-285, 288
— alucinación de la, 274
— *buffer* de, 159n., 231
— como biblioteca, 135, 145
— de acceso aleatorio (RAM), 226
— defectuosa, 330
— de los ordenadores, distinta de la del cerebro, 233, 238
— de trabajo (*véase también* espacio de trabajo), 278-280
— ecoica, 159n.
— episódica, 291
— pérdida de (*véase también* amnesia), 263
— semántica, 347 n.
— y color, 385
memosfera, 219, 233
mental, imagen (*véase* imagen mental) Cielo de las Imágenes Mentales, el espacio fenoménico como, 144n.
mentales (*véase también* lenguaje del pensamiento), 245, 248, 376
mentalidad ágrafa, 234
mentalistas, términos, 50-51
mente
— como nido de memas, 219
— como regularidad percibida por una mente, 322
— es el cerebro, 44
— mente/cerebro, 50
— oído de la mente, 70, 71
— ojo de la mente, 65, 309, 311, 323, 361, 366-368, 383, 385, 400
— sustancia mental, 44-47, 48, 98, 355
Menzel, E., 438, 493
Merikle, P., 139, 496
meta-conocimiento, 354, 449
metacontraste, 155-156
metáfora, 96, 103-106, 144, 194, 244, 288, 300, 324-326, 466
Metzler, J., 299, 498
microhábito, 230, 232, 268
milagro, 50, 253, 269, 466
milisegundo, 117
Miller, G., 194
Minsky, M., 14, 74n., 121, 253, 255, 272, 274-276, 286, 308n., 326n., 367, 370-371, 437, 449, 493
Mishkin, M., 347n., 493
Misterial, Nuevo, 287
misterio, 13, 49, 53, 322
— definición, 33
— de la conciencia, 27, 28-29, 34, 35, 37
— enfrentado, 294
— sistemático, 417
mito, 464-465
Modelo de las Versiones Múltiples de la conciencia, 29, 125-127, 183, 241, 267-268, 271, 277, 346n., 369, 382, 404, 409, 419, 443, 466, 468
modularidad, 274
Monod, J., 187, 493
mono
— preferencias de colores del, 396
— verde, 207
Monroe, M., *véase* Marilyn (Monroe)
moralidad, su relación con la conciencia, 459-466
Morse, código, 62, 353n.
mostrar frente a referir, 309-311, 329
Mountcastle, V., 275, 493
movimiento, aparente, 147, 163-164, 349
— detección del, 371-372
movimiento percibido frente a movimiento inferido, 137
Moynihan, D., 368
MT (córtex medio-temporal), 142n.
múltiple, función, 286, 290
mundo (*véase también* heterofenomenológico, mundo) ficticio, 144
— heterofenomenológico como ficticio, 93
murciélago, 452-460

Nabokov, V., 312, 494
Nagel, T., 84, 109, 287, 384, 435, 436n., 445, 452-454, 459, 467, 471, 494
narcisismo del diseño cerebral, 394
narrabilidad, 348
narración, 428-429
— creación de una, 258
— de un sueño, 25-26
— el observador como un ovillo de, 150
— explicación evolutiva como, 185-186
— flashbacks en una, 162
— flujo narrativo, 127
— fragmentos narrativos, 149, 268, 273
— marco narrativo, 450
— orden de composición, 162
— precipitados por sondeos, 148-150
— secuencia narrativa, 149
navegación, 115, 164-165
Necker, cubo de, 302
negligencia, 366-370
— ante los errores tipográficos (*véase también* el efecto del lector de pruebas), 368
— benigna, 368, 373
— financiera, 369
Neisser, U., 24, 159n., 293, 494
Nemirov, L., 410, 494
Neumann, O., 14, 193, 202, 235, 367, 494
neuroanatomía, 284n., 287
neurociencia, 28, 40, 51, 84, 161, 269, 271, 276, 282, 388, 394
— principio guía de la, 83

neurocirugía, 70n., 168, 181
— neurocirujanos malvados, 402, 403
Neurofilosofía, 173
Neurona lógica, 230n.
Neurona Pontificia, 242, 424
neutralidad, 85, 98
— de la heterofenomenología, 84, 96, 107-111
neutrino, 48
Newell, A., 159n., 278, 280, 283, 284n., 289n., 494
Nielsen, T., 126n., 494
Nietzsche, F., 241
Nilsson, N., 14, 99, 494
Nisbett, R., 197n., 489
nivel de baudios, 253, 369
nivel de detalle, 105
nivel de explicación, 289-291
nivel del conocimiento, 290n.
NMDA, receptor, 287
noemas, 56
Norman, D., 202, 288, 291, 494
Nottebohm, F., 196, 494
noúmenos, 56
novela aleatoria, 17n.
Nowlan, S., 199, 488
Nueva Orleans, Batalla de, 160, 163, 182-183

Oakley, K., 434, 494
objetivos, 279
observación
— conocimiento sin, 328n.
— interna, 332
observador, 115, 116, 143, 150, 244, 265, 269-270, 416
— en el cerebro, 120, 180, 325, 327, 443
— localización del, 121, 138
— su papel en la definición de ciertas propiedades, 391-393
observadores de pájaros, 347
ocurrente, propiedad, 382-384, 397-398, 400
odio a las serpientes por parte de los primates, 397-398
oído, 58-64
ojo interno, 329
olfato, 58
ontología, 48
operación aritmética, 227
operacionalismo, 109, 140
— definición, 130n.
— de la primera persona, 145, 146-147
opinión, opuesto a creencia, 90n.
optimalidad, supuesto de
ordenador, 159-160
— asíncrono, 160
— que habla, 88, 91, 104-107, 224-233, 243, 446-451
organización temporal
— absoluta, 175-176, 178-180, 182
— de los acontecimientos cerebrales, 127, 167, 178-179
— de los acontecimientos conscientes, 127, 137-138
— qué determina su contenido, 163-164
órgano del teatro de la mente, 61
orientación, respuesta de, 193, 201-203, 236
Ornstein, R., 382, 495
orwelliano frente a estaliniano, 130-138, 139-142, 146-147, 156, 173-176, 177-178, 181-182, 331, 346n., 406, 452, 471
Otto, 243-244, 250, 255, 316-318, 322, 329-330, 332, 344, 349, 369-371, 373-379, 386, 394-395, 398-400, 413-415, 467
Oxford, 472

paciente con el cerebro dividido, 211, 273, 434-436
Pagels, H., 177, 495
paisaje adaptativo, 198-201
Palacios, A., 361n., 387, 499
Pandemónium, 202, 235, 250-252, 254-256, 264, 267, 275, 317, 328, 347, 466, 475-476
Papá Noel, 97, 144n.
papel pintado, 365-366, 370-371, 375
— experimentos con, 475-477
Papert, S., 230n., 495
paradoja, 28
— de la conciencia, 115
paralelo
— arquitectura, 223, 280
— en el cerebro, 124-125, 127, 129, 147-149, 222, 228-229
— procesamiento, 202-203, 230, 251-252, 256-257, 264, 267, 272-273, 282-283
— procesamiento distribuido (*véase también* conexionismo), 189n., 281-283
— simulado, 231, 279-280
paranormal, 337, 430
parásito, 425
— los memas como, 218, 235, 268
Parfit, D., 433, 495
parlamentario, 285-286, 368
Partee, B., 237
Pascal, B., 315
patología, 261-265, 287-289, 333-334, 347n., 367-370
Pavlov, I., 134
PC-Paintbrush, 358-360
Pears, D., 265, 495
pena de muerte, 465
Penfield, W., 70n., 168, 495
Penrose, R., 48-49, 116n., 169, 441
pensamiento, 28, 330
— como hablar a uno mismo, 70
— de orden superior, 327-328, 331, 343-344
— inconsciente, 321-322
— opuesto a creencia, 320, 321
— pensar con pensamientos, 311
— sin imágenes, 71
Penseur, Le, 236
Pequeño Esbozo, 267-268, 269-277
Perceptrón, 230n.
perla mental, 378-379
Perlis, D., 291, 437, 495
personaje ficticio, el yo como, 421, 439
Petrone, P., 369, 481
phi, fenómeno, 128, 474
piano, afinar un, 348
Piantanida, T., 82n.
Pinel, P., 120n.
Pinker, S., 213n., 308n., 495
pista. *Véase* ayuda curiosidad, 27, 223
Pitts, W., 230n., 493
píxel, 286, 310, 360, 364
plan, 192, 354
— oportunista, 256
plasticidad
— de los ordenadores, 223-224
— del sistema nervioso, 195

— fenotípica, 199-201, 206, 222, 233
Platón, 236
— aviario de Platón, 235-236, 238, 279, 284, 293, 314
— memas platónicos, 218-219
plazo, 164-167
pleno, la conciencia como un, 377, 418-419, 476-477
poderes, y no las funciones, de los mecanismos de la conciencia, 290
Pollatsek, A., 347n., 373, 495
polo auditivo del barbero, 81
poltergeist, 47
poner entre paréntesis, 56
Pöppel, E., 120, 495
Popper, K., 41, 45n., 168-169, 173, 495
porque sí, respuesta, 207
positivismo, 402
Potemkin, 21
potencial evocado primario, 171-172, 176
Potter, B., 436
pre-experiencial frente a post-experiencial, 122, 132-133, 136-137, 142n., 260, 405-407
precognición
— desechada, 129
— por el carrusel, 180-181, 475
preconsciente, 261n.
Premio Nobel, 52, 269
presentación (*véase también* representación), 121, 146-147, 167, 182-183, 268, 270, 325-326, 375, 405, 456
presentimiento, 354, 375-376, 467
preverbal, mensaje, 248-249, 252, 254, 260
primera persona, 79-80
— idoneidad, 214, 216, 218
— operacionalismo de la, 146
— perspectiva (o punto de vista) de la, 79-80, 82-83, 136-137, 138, 347
primitiva (operación de un ordenador), 224-227, 249, 227-228
principio del mínimo esfuerzo, 253n.
privilegiado, acceso, 76, 81, 136, 259-260
— no existe en los procesos de producción del habla, 109-110, 323
problema del ligamiento, 133, 271
problema epistemológico de las otras mentes, 453, 458
problema primordial, 190-191
procesador de textos, 324-325
— el cerebro como, 239
— el ordenador no concebido como, 225
procesamiento central (*véase también* CPU), 274
procesos caóticos, 196
profundidad, percepción de la, 125
programación de conflictos, 202
programa
— como descripción o lista de instrucciones, 229
— como lista de instrucciones, 278
— como receta, 230
— de la máquina virtual del cerebro, 232
Prolog, 315
propiedad intrínseca, 75-77, 383-385, 395, 397, 399, 408, 420, 440, 446, 460-462, 467
propiedad
— disposicional, 383-385, 387, 391
— ocurrente, 383-384, 388
propiocepción, 16-17, 57-59
proposición, 375-376
prosopagnosia, 82
protocolos conscientes, 340-341, 343
protocolo, base para un, 346, 348
proyección, 140, 161n.
— en el espacio, 142, 144, 145
— hacia atrás en el tiempo, 142, 143, 144, 153
proyector en carrusel, 180-181, 475
Psicoanálisis (el juego), 22-27, 252
psicofísica, 394
psicología, 51, 86-87, 268-270, 308, 412
— del sentido común, 316, 319, 322, 327-332
publicación, como una mala metáfora de la conciencia, 139-140
punto ciego, 334-335, 355-356, 360, 366-368
punto de vista, 115-116, 139
— cambio de, 352
— creado por replicación, 188, 190
— del observador, 190
— esparcido, 121, 149, 166
Putnam, H., 293, 402, 470, 495
Pylyshyn, Z., 162n., 284n., 495

qualia, 29, 355, 369-370, 381-422, 445-446, 461, 466, 469-470
— definición, 76-77, 349
qualófilo, 398, 399, 401-402, 403-406
Quine, W., 393n., 495

Raab, D., 155, 486
racionalidad, 265
— ideal, 314
— mínima, 293n., 315
— presunción de, 468
radar, 437-439
Ramachandran, V., 273, 276n., 364n., 476, 495
Ramberg, B., 14
ramificaci'on condicional, 226, 279
Ramsey, W., 282, 496
RAM. *Véase* memoria de acceso aleatorio
Raphael, B., 99, 496
rastreador de ojos, 372-373, 476-477
rastrear, 191, 204, 344-346
Rayner, K., 347n., 373, 495
razonamiento práctico, 264-265
razones, nacimiento de las, 187-188, 190, 424-425
re-presentación, 127, 306-307, 355
Reaganomics, 123
realidad de los objetos intencionales, 96
realimentación, 259-261, 345, 346, 348
— cerrando el bucle, 342
realismo, 469-471
reconocimiento de patrones, 234, 279, 283, 307, 387
recordar, 210, 291, 309, 353
recursividad, 269, 323
red Quien Gana se lo Queda Todo, 202
Reddy, R., 278, 496
reduccionismo, 75, 398, 465-466, 469-470
reentrada, circuitos de mapas de, 189n., 281
referencia, teoría causal de la, 285n.
«referencia hacia atrás en el tiempo», 168, 171-172
referir, 243-245, 252, 265, 270, 330
— frente a expresar, 167, 316-317, 327-328, 331
reflectancia espectral de las superficies, 387
reflejo, 44, 115, 135
reflexión, 332, 341, 459
refuerzo, 234

registro de instruccines, 227, 239
registro, 232
— de memoria, 227
— números en un, utilizados para representar, 361
regresión infinita
— cómo detenerla, 330
— de Significadores, 253
regularidad, precibida desde la actitud intencional, 468
Reingold, E., 139, 496
Reisberg, D., 304, 308, 496
relatar historias, 271, 314, 428, 438-439
relativismo, 392
«repleción», 136-137, 140-142, 172, 355-367, 376
replicador, 187-189, 213-216, 218-219
representación, 204-206
— de uno mismo, 427-429, 439-440
— en el córtex, 284
— mínima, 191-192
— por parte de un sistema inmunológico, 188
representación toscamente continua, 360, 364, 365n.
representado frente a representante, 93-94, 144, 150, 157-158, 161-163, 164-165, 175, 177-178, 180
reproducir, 305-307, 310, 366, 375, 387
res cogitans, 41, 42, 50, 53, 119
resolución de conflictos, 231, 280
— en el córtex prefrontal, 190
responsabilidad moral, 245
Respuesta de los Sistemas, 451
respuesta galvánica cutánea, 83, 342n.
Richard, R., 197n., 431, 496
risa, 74-77
Ristau, C., 457, 496
Rizzolati, G., 201, 496
robot, 443, 460
— animales, 55, 463
— consciente, 443, 446
— replicador, 188
Rodin, A., 236
Roelofs, C., 141
Roepke, M., 14
Romeo y Julieta, 93
Rorty, R., 80, 471, 496
Rosenberg, J., 14, 146n.
Rosenberg, Julius y Ethel, 388, 394, 412
Rosenbloom, P., 159n., 278, 280, 283, 284n., 496
Rosenthal, D., 14, 318, 320-321, 324, 326-327, 330, 344, 496
Rozin, P., 272, 396, 425, 496
Rudolph the Red-Nosed Reindeer, 399
Rueda de Oraciones Tibetana, 295-296
Rumelhart, D., 234, 253, 282, 493, 497
rumores, como modelo de acceso, 329
Russell, B., 143n., 259, 275
Ryle, G., 45, 61, 236, 424, 497

sacádico, movimiento, 66, 125, 194, 346, 365-366, 371-373, 475-477
— supresión sacádica, 372-373
Sacks, O., 314, 334n., 459, 497
Sandeval, E., 272, 497
Sanford, D., 431, 497
Sangree, M., 14
Sartre, J.-P., 177, 497
satisfacción de condiciones, 253
Sato, S., 347n., 500
Savage-Rumbaugh, S., 397n., 493
Schacter, D., 366, 369
Schank, R., 271, 449, 497
Scheerer, E., 14
Schossberger, C., 14
Schull, J., 197, 497
Searle, J., 109, 283, 292, 333-334, 447-452, 467, 470, 497
secreto, 458-459
secuencia subjetiva, 150, 168, 180, 183,
secundaria, cualidad (*véase* cualidad, primaria y secundaria)
Segunda Ley de la Termodinámica, 218
Selfridge, O., 14, 202-203, 498
Sellars, W., 77, 383, 395, 498
semántica de líneas, 100-101
semántico/-a
— análisis, 87-89
— disponibilidad, 148
— niveles de simularidad, 93
semiótico, materialismo, 421
sensaciones, 29, 64
«sensa», 384
sentimientos en bruto, 334, 384
sentir, 349
serial
— búsqueda, 292
— encadenamiento, 235
— proceso, 226, 228, 232, 249, 253, 254, 265, 268, 272-273, 277
serie transformada de Fourier, 60
sexo, 91
— evolución del, 186-187
Seyfarth, R., 207, 455, 457, 481
Shakespeare, W., 93
Shakey el robot, 99-107, 109, 144, 192n., 243-245, 300, 305, 323-324
— como zombi, 322
Shallice, T., 202, 269, 288, 291, 334n., 494, 498
Sharpe, T., 424, 498
Shepard, R., 82n., 107, 299-303, 305, 365n., 498
Sherman, V., 196, 493
Sherrick, C., 156-157, 487
Shipp, S., 148, 502
Shoemaker, S., 381, 395, 402, 498
SHRDLU, 104-105
Shylock, 417
SI-ENTONCES, primitiva (*véase también* ramificación condicional), 278-279
Siegel, R., 25, 498
significado, 87-90, 247, 248, 253, 257-259
— y proto-significado, 192
Significador, 285, 259, 261
— Central, 242, 244, 247, 252, 259, 260, 265, 267, 271, 316, 375-378, 432, 443, 466
silicio frente a orgánica, máquina, 42-43
símbolos en movimiento, 284n.
simetría vertical, detección de la, 192, 201
Simon, H., 87, 307, 486, 491, 498
simpatía, 73
simultaneidad, subjetiva, 121, 149, 177, 178
síndrome de Anton, 91, 369-370, 479
síndrome de Korsakoff, 263
Síndrome de los Filósofos, 412
singularidad, 116n.

sistema de producción, 279-281
sistema de símbolos físicos, 290n.
sistema experto, 42
sistemas especialistas en el cerebro, 194, 202, 208-209, 236, 242, 254, 267-268, 271-273, 277, 288, 290-291, 362-363, 371
— en tanto generalistas, 383-385
Skinner, B., 197
Sloman, A., 14
Smith, W., 434, 490
Smith, Y., 339n., 481
Smolensky, P., 282, 498
Smullyan, R., 146, 498
Smythies, J., 143, 145, 173, 499
Snyder, D., 150n., 499
Soar, 278, 281-282, 285, 293
software, 232-234, 466
— definición, 224
— en el cerebro, 203, 223
— nivel del, 445-446, 451
solipsismo, 414
sonambulismo, 43-44
sondeo, 127, 149, 157, 183
— momento no óptimo para efectuarlo, 149
sonido, velocidad del, 116, 120
sonido en estéreo, experiencia del, 143
soñar, 71
sordo-mudo, 459
sospechosas, cualidades, 391-392
Souther, J., 347n., 500
Spalding, O., 14
Sperber, D., 14, 208, 253n., 499
Sperling, G., 159n., 499
Sperry, R., 292, 499
Spillman, L., 82n., 499
Spinoza, B., 201, 499
Stafford, S., 14, 44, 499
Stalnaker, R., 293n., 499
Stanford Research Institute, 99
Steinberg, L., 14
Stich, S., 282, 496
Stoerig, P., 337, 499
Stoll, C., 140, 146, 499
Straight, H., 311, 499
Stratton, G., 404, 499
Strawson, G., 470, 499
Strawson, P., 43, 499
Studdert-Kennedy, M., 62, 491
subjetividad, 146, 173, 383-384, 401
subjetivismo, 392
substantia nigra, 382
sueño, 22-23, 25-27, 72, 253
sufrimiento, 460-464
sujeto
— heterofenomenológico, 141, 144
— unificado, 90-91
Sullivan, A., 220
superficie de la mente consciente, 237, 326n.
Supermán, 37
suprimir algo, 268-269

TAC (tomografía asistida por ordenador) (CT), 336
tacto, 58
— mantener el contacto, 345
tálamo, 53, 287
taquitoscopio, 155
Taylor, D., 402, 499
tejido de discursos, 421, 427
tejido milagroso, 52
teletransporte, 440
televisión, 18
— velocidad de la, 117
teorías de reflector de la atención, 287-288
tercera persona, perspectiva (o punto de vista) de la, 82-84, 108-109, 141-142, 347, 453-454, 456-457
termita, 426-427
terremoto, 35-36
Test de Turing, 323-324, 447-451
Testigo, 333-334, 369, 432, 443, 466
— la sustancia mental como, 40-41
testimonio de los sentidos, 329
tetracromática, visión, 361n.
Thagard, P., 197n., 489
Thompson, D'A., 185
Thompson, E., 14, 361n., 382, 387, 389, 499
Titchener, W., 56
tiempo
— cómo se representa en el cerebro, 158-167
— de la punta de los dedos, 176
— prisa, 158, 164-165
— real, 141, 163
— ventana temporal, 133, 158, 165-166
toma de decisiones, 178, 254
toma doble, 293
tono armónico, 61
TPF (Trastornos de Personalidad Fraccionada), 433-434
TPM (Trastornos de Personalidad Múltiple), 430, 432-433, 434
Trampa de la Introspección, 371
Tranel, D., 82n., 499
transcripción de los sonidos del habla, 87-88
transmisión en el cerebro, 209, 271, 292
TRC (tubo de rayos catódicos), 224, 229, 300, 303, 305, 308, 310, 316, 476
Treisman, A., 347n., 500
Trevarthen, C., 434, 491
tricromática, visión en color, 361n., 389
troceado, 281
truco
— aprender un nuevo, 203
— Buen Truco, 197-200, 201, 210, 212-213, 221, 238, 390
— conjurarlo (*véase también* magia; juego de manos), 140, 472
— en el diseño biológico, 44
— hacer trampas con el tiempo, 169
— la heurística como un maletín de trucos, 293
— las alucinaciones como bromas gastadas por el cerebro, 20
Truman, H., 44, 440
Turing, A., 223-226, 227, 228, 231, 249, 277, 279, 323
Turing, máquina de, 224-225, 279
— Universal, 223, 228
Tye, M., 410, 500

Ungerledier, L., 347n., 493
urdidores de slógans, 470
Usuario de la biblioteca (*véase también* ilusión, del usuario), 371
Uttal, W., 161n., 500

Valéry, P., 190, 433
Vallar, G., 488
van der Heiden, L., 14
Van der Waals, H., 141, 500
Van Essen, D., 148, 500
Van Gulick, R., 14, 339, 410, 500
Van Hoesen, G., 82n., 482
Van Tuijl, H., 361, 500
Van Voorhis, A., 14
Varela, F., 361n., 387, 499
vegetarianismo, 465
Velleman, J., 470, 480
velocidad de cómputo, 231n., 232
Vendler, Z., 45n., 500
ventana
— de control, 183
— de edición, 173
— temporal, 133, 158, 164-166
verificacionismo, 140, 146, 401, 414, 470-471
ver, 345-347, 349, 366
— y comprender, 67
Vesalius, 118
Vesubio, 43-44
veto por parte del yo, 178, 182
Victoria, La Reina, 94
videojuego, 17-18
vídeo, cinta de, 360
virtual
— capitán, 242
— conexión, 209
— espacio, 302, 305
— estructura, 314
— lenguaje de programación como máquina virtual, 315-316
— máquina, 223, 223-224, 228-234, 238-239, 242, 238-239, 242, 267-268, 272-274, 283, 295, 324-325, 443, 445-447, 449, 458, 466
— objeto, 301
— presencia, 371-373
— Realidad, 18n.
virtus dormitiva, 75-76, 397, 440
visión, 63-68, 292-293, 334-367
— localizada en el interior de los ojos, 122
— parafoveal o periférica, 66, 81, 365, 371, 373, 475-476
— por ordenador, 101-105
— resolución de la, 58
visión ciega, 261, 334-344, 348-350, 354, 369-370, 451, 476
— en los murciélagos, 456
— entrenamiento en la, 342
visión protésica, 349-353
visualización, 308-309, 377
vitalismo, 37, 396
Vivaldi, 38
von der Malsburg, C., 287, 500
von Grünau, M., 128, 133-135, 491
von neumanniano, 223, 227, 238, 272
von Neumann, J., 223-224, 227, 228, 230n., 278
von Uexküll, 457, 500
Vorsetzer, 306, 308, 310, 325, 386
Vosberg, R., 25, 500

Waller, F., 249
Walton, K., 91, 377, 439, 500
Warhol, A., 365
Warren, R., 500
Warrington, E., 262, 491
Wasserman, G., 178-179, 500
Waterhouse, L., 273
Weiner, P., 14
Weinstein, S., 14
Weiskrantz, L., 337, 339n., 342n., 344, 349, 501
Welch, R., 404, 501
Werner, J., 82n., 499
Wertheimer, M., 128, 501
West, L., 25, 498
West Side Story, 93
White, S., 381, 501
Whiten, A., 457, 481, 501
Wiener, N., 190, 501
Wiesel, T., 52
Wilkes, K., 452, 456, 501
Wilson, D., 208, 253n., 499
Wilsson, L., 426, 501
Winograd, T., 104, 501
Wittgenstein, L., 69n., 76, 355, 402, 408n., 458, 471-472
Wolfe, J., 291, 501
Woodfield, A., 14
WordPerfect, 232
WordStar, 229, 232
Wordsworth, W., 33-34
Wundt, W., 56
Wyeth, 396
Wynes, K., 14

«¿Y ahora qué?», 269, 277-289
Yo Interno, 317
Yonas, A., 192, 501
Young, J., 197n., 501
«Yo», 436-437
yo, 29, 41, 177, 213, 242, 292, 420, 423-441
— biológico, 424-429, 437, 458
— como algo no independiente de los memas que alberga, 221, 315-316
— como alma, 379
— localización del, 177-178

Zajonc, R., 413, 501
Zeki, S., 148, 502
Zihl, J., 344-345, 502
zimbo, 323-324
zombista, 269
zombi, 85-86, 88, 91, 96, 107, 295-296, 315, 322-324, 326, 416-417, 461, 471
— el murciélago como, 455-456
— sujeto con visión ciega como un parcial, 334-345
zoología, 55-57, 79-82

La conciencia explicada
Daniel Dennett

Paidós
Básica

Detallada investigación sobre la conciencia humana, el presente libro, uno de los más originales ensayos de los últimos tiempos, refuta la teoría tradicional de la conciencia basada en el sentido común y presenta un nuevo modelo fundamentado en una ingente cantidad de datos, tomados a su vez de campos tan diversos como la neurociencia, la psicología, la filosofía y la inteligencia artificial. La visión de Dennett sobre el tema, así, es antiintuitiva pero convincente, y, como en toda teoría revolucionaria, su fuerza y su capacidad para sorprender aparecen íntimamente relacionadas. El resultado es un útil y accesible recorrido por el campo de la filosofía de la mente a partir del cual nuestras actuales teorías sobre la vida consciente —de las personas, de los animales e incluso de los robots— puede que se vean profundamente afectadas por nuevas y fascinantes perspectivas. Sin duda uno de los mejores ejemplos de libro científico dirigido tanto a profesionales como a lectores no especializados, *La conciencia explicada* constituye finalmente un trabajo filosófico de la mejor especie: sencillo, pero sin caer en la trivialidad; bien documentado, pero sin abrumar con un exceso de información. Como ha dicho Howard Gardner, "una notable reflexión sobre la conciencia... a cargo de uno de nuestros más sobresalientes pensadores". Daniel Dennett es catedrático de Humanidades y Ciencias, director del Center for Cognitive Studies de la Universidad Tufts y autor de títulos tan celebrados como *Brainstorms: Philosophical essays on Mind and Psychology* y *The Mind's I*, este último junto con Douglas Hofstadter.

ISBN 84-493-0170-X
32076
9 788449 301704

Diseño: Mario Eskenazi